PRAGMATICS OF UNCOOPERATIVE AND MANIPULATIVE COMMUNICATION

Thèse de Doctorat présentée à la Faculté des Lettres et Sciences Humaines

Institut des Sciences du Langage et de la Communication

Université de Neuchâtel

Pour l'obtention du grade de Docteur ès Sciences Humaines

par

Steve OSWALD

Thèse de doctorat dirigée par le professeur Louis de SAUSSURE (Université de Neuchâtel)

Membres du Jury:

Professeur Paul CHILTON (University of Lancaster)

Professeur Fabrice CLÉMENT (Université de Neuchâtel)

Professeur Frans van EEMEREN (University of Amsterdam)

Professeur Didier MAILLAT (Université de Fribourg)

Soutenue le lundi 13 décembre 2010

UNIVERSITÉ DE
NEUCHÂTEL

Faculté des Lettres et Sciences Humaines
Espace Louis-Agassiz 1
CH - 2000 Neuchâtel
Tél : +41 (0)32 718 17 00
Fax : +41 (0)32 718 17 71
secretariat.lettres@unine.ch

IMPRIMATUR

La Faculté des lettres et sciences humaines de l'Université de Neuchâtel, sur les rapports de M. Louis de Saussure, directeur de thèse, professeur à l'Institut des sciences du langage et de la communication de l'Université de Neuchâtel; M. Frans van Eemeren, professeur au Department of Speech Communication, Argumentation Theory and Rhetoric à l'Université d'Amsterdam ; M. Fabrice Clément, professeur à professeur à l'Institut des sciences du langage et de la communication de l'Université de Neuchâtel ; M. Didier Maillat, professeur associé au Département d'anglais à l'Université de Fribourg ; M. Paul Chilton, professeur au Department of Linguistics and English Language de l'Université de Lancaster (UK), autorise l'impression de la thèse présentée par M. Steve Oswald en laissant à l'auteur la responsabilité des opinions énoncées.

Neuchâtel, le 13 décembre 2010

Le doyen
Laurent Tissot

Acknowledgements

I would first like to express my deep gratitude to Louis de Saussure. By accepting to supervise my work from my very first steps in Linguistics and Pragmatics, by being one of the first to envisage manipulation through a cognitive pragmatic lens, but also by providing me with unfailing guidance, support and encouragement, he has brought together the necessary conditions for the completion of this thesis. I feel fortunate to be one of his students.

I want to thank my colleagues at the University of Neuchâtel and in particular the members of the Semantics and Pragmatics Research Group, Patrick Morency, Alain Rihs, Cécile Barbet, Thierry Raeber and Andriy Sytnyk, who have always generated a highly stimulating work environment through valuable scientific discussion. I was lucky enough to collaborate very closely in particular with Patrick and Alain on specific projects and wish to thank them for both their friendship and scientific collaboration.

I thank the Swiss National Science Foundation for awarding me a grant for prospective researchers which allowed me to spend a year at Lancaster University at the Department of Linguistics and English Language in 2008-2009. This was a rewarding human and professional experience and I wish to thank the members of the LIP Research Group, in particular Majid KhosraviNik, Alexandra Polyzou, Costas Gabrielatos, Adèle Petitclerc and Paul Sarazin, who always welcomed a critical discussion on the proper ways of going about in the study of discourse and communication. The input I garnered over these exchanges has been of great value to some discussions in Chapter II of this thesis.

I am very grateful to the members of the jury, Paul Chilton, Fabrice Clément, Frans van Eemeren and Didier Maillat, who have agreed to evaluate this work. I had the opportunity to interact with all four of them over the last years and would like to thank them here for taking the time to discuss their scientific work with me, what is more always with commitment to great clarity and thoroughness. The research they have conducted over the years in their respective fields has been truly inspirational and I would like to acknowledge the significant influence they have had on my research. The CSC model presented here would furthermore not exist today if Didier Maillat had not accepted to collaborate with me on its elaboration in the first place. I particularly wish to thank him for his ever-stimulating collaboration on this scientific project we embarked on four years ago.

I would also like to thank scholars and colleagues whom I had the opportunity to have stimulating scientific exchanges with: Salvatore Attardo, Joana Garmendia, Chris Hart, Thierry Herman, Constanza Ihnen, Scott Jacobs, Pierre Larrivée, Marcin Lewinski, Frédéric Minner, Andrea Rocci, Francisca Snoeck Henkemans, Danièle Torck and Ken Turner have all contributed, through discussion, to push my reflection further.

Outside Academia, there are extremely important people I want to thank, because they have always believed in me and have provided me, both materially and emotionally, with everything I needed to carry on the doctoral track. My parents, Eric and Francine, who have always believed in me – it would never have been possible for me to go to University were it not for their unfailing support; my sister Jenny and my nieces, Tatiana and Alexandra, my aunt Dolly, my

Granddad Arthur, and my late Grandmothers Alice and Claudine. My gratitude also goes to my closest friends, Ralph, Carlos, Christophe, Santiago, Patrick, Frédéric, Conrado, Sylvestre, Mehdi and Jérémie, who are still here after so many years and who have been there for so long.

Key words: cooperation, uncooperativeness, manipulation, manipulative communication, deception, argumentation, persuasion, pragmatics, cognition, contextual selection constraint, relevance, salience, fallacy, bias.

Abstract: This dissertation tackles the issue of uncooperative and manipulative communication from a cognitive pragmatic perspective and develops and documents the fundamental hypothesis that its success depends on whether the manipulator manages to keep the manipulative intention concealed. Specifically, manipulative communication is conceived here as a twofold process meant to i) mislead the addressee into processing limited sets of contextual information and ii) prevent her/him from processing contextual information that would be detrimental to the success of the manipulative attempt.

The dissertation draws on several fields of research in the Humanities and attempts to interface findings from cognitive psychology (mainly research on cognitive biases) and argumentation theory (research on fallacies) into a consistent cognitive pragmatic account of information-processing. To this end, the so-called Contextual Selection Constraint model (CSC) is presented; this model specifies from a theoretical perspective how certain linguistic and argumentative strategies can be used to constrain the comprehension procedure so that targeted assumptions end up partaking in the derivation of meaning and other unwanted assumptions turn out to be disregarded – or unprocessed altogether. These possibilities are conceived as natural potential consequences of our cognitive system's inherent fallibility.

Mots clés: coopération, non-coopération, manipulation, communication manipulatoire, tromperie, argumentation, persuasion, pragmatique, cognition, contrainte de sélection contextuelle, pertinence, saillance, sophisme, biais.

Résumé: La présente étude traite de la communication non-coopérative et manipulatoire d'un point de vue pragmatique cognitif; elle développe et documente en outre l'hypothèse fondamentale selon laquelle le succès de ce genre de communication repose sur la capacité du manipulateur à dissimuler son intention manipulatoire. En ce sens, la communication manipulatoire est ici définie comme un double processus: d'une part, elle vise à amener un destinataire à se représenter un ensemble restreint d'informations contextuelles, et d'autre part elle tente simultanément de l'empêcher de se représenter un autre ensemble d'informations contextuelles qui seraient nuisibles au succès de la tentative de manipulation.

Ce travail s'inspire de divers champs de recherche en Sciences Humaines et propose une interface entre la psychologie cognitive (principalement la recherche sur les biais cognitifs) et les théories de l'argumentation (l'étude des sophismes en particulier) au sein d'un cadre théorique pragmatique cognitif capable d'héberger une telle intégration. Dans cette optique, le modèle de la Contrainte de Sélection Contextuelle (CSC) est présenté; celui-ci décrit et explique d'un point de vue théorique le fonctionnement de certaines stratégies linguistiques et argumentatives susceptibles de contraindre la procédure de compréhension. Notamment, il s'attache à montrer comment il est possible de favoriser l'inclusion de certaines représentations contextuelles dans le calcul du sens, mais aussi de quelle façon il est tout autant possible d'empêcher d'autres représentations de faire partie de ce dernier. Ces possibilités sont conçues comme des conséquences naturelles, mais non moins potentielles, de la faillibilité inhérente de notre système cognitif.

CONTENTS

INTRODUCTION **1**

CHAPTER I. COOPERATION, NON-COOPERATION AND MANIPULATION: THEORETICAL OVERVIEW **9**

I.0. INTRODUCTION 10

I.1. COOPERATION 10
I.1.0. Overview *10*
I.1.1. Grice's rationale for cooperation *11*
I.1.1.1. General remarks 11
I.1.1.2. Implications of cooperation 15
I.1.2. Cooperation for communication simpliciter: communicative cooperation *16*
I.1.2.1. The cooperative paradox and the dichotomy of the CP 16
I.1.2.2. Communicative cooperation 19
I.1.2.3. Communicative cooperation as rationality of communication 23
I.1.3. Informative cooperation *27*
I.1.4. Perlocutionary cooperation *31*
I.1.5. Evolutionary bases of cooperation. *35*
I.1.5.1. The prisoner's dilemma and reciprocal altruism 36
I.1.5.2. Communication, language and cooperation in evolution 40
I.1.6. Summing up: the cooperative status of locution, illocution and perlocution *45*
I.1.7. Post-Griceans: Relevance Theory *48*
I.1.7.1. Why the CP does not explain comprehension and communication 49
I.1.7.2. Allott: coordination vs. cooperation 50
I.1.8. The descriptive value of cooperation *54*
I.1.9. Conclusion *58*

I.2. NON-COOPERATION 61
I.2.0. Introduction *61*
I.2.1. Grice on non-cooperation *62*
I.2.1.1. Maxim violation and non-cooperation: preliminary remark 62
I.2.1.2. Ways of failing to fulfil conversational maxims 65
I.2.1.2.1. Unostentatiously violating a maxim 66
I.2.1.2.2. Opting out 68
I.2.1.2.2.1. Opting out of the operation of conversational maxims 68
I.2.1.2.2.2. Opting out of the CP 72
I.2.1.2.2.3. Opting out as a process vs. opting out as a state of affairs 73
I.2.1.2.2.4. Opting out, overtness and cooperation 76
I.2.1.2.3. Clash between maxims 77
I.2.1.2.4. Flouting 79
I.2.2. Intermediate summary: overt uncooperativeness *81*
I.2.3. Deception as unostentatious violation of conversational maxims *85*
I.2.3.1. Deception and Information Manipulation Theory (IMT) 86
I.2.3.2. Limitations of IMT 88
I.2.3.2.1. Problems with IMT: missing links between data and theory 89

I.2.3.2.2. IMT fails to understand the Gricean framework 93
I.2.3.2.3. Problems with identifying and ascertaining deception 94
I.2.4. Summing up: CC, IC, PC and deception 96

I.3. MANIPULATION 103
I.3.1. Definitional criteria of manipulation: a survey 103
I.3.1.1. Truth 103
I.3.1.1.1. Lies 104
I.3.1.1.2. Manipulation and lies 109
I.3.1.2. Speaker interest 111
I.3.1.3. Covertness 114
I.3.1.4. Social inequality 116
I.3.1.5. Intention 119
I.3.2. Manipulation as non-cooperation: towards a working definition 122
I.3.2.1. A descriptive definition 123
I.3.2.2. Explanatory issues: specifying the working definition of manipulation 125
I.3.3. Neighbouring phenomena 130
I.3.3.1. Persuasion vs. manipulation 130
I.3.3.2. Deception vs. manipulation 133
I.3.3.3. Argumentation vs. manipulation 135
I.3.4. Conclusion 136

CHAPTER II. A CRITICAL VIEW ON THE MAIN APPROACHES TO MANIPULATIVE COMMUNICATION, FALLACIES AND DECEPTION 139

II.0. INTRODUCTION 140

II.1. CONSTRAINTS ON DISCOURSE 141
II.1.1.Critical Discourse Analysis 141
II.1.1.1. The critical stance: historical roots of CDA 141
II.1.1.2. On the scientificity of CDA: theory, multidisciplinarity and empirical research 145
II.1.1.2.1. The status of CDA as a scientific theory 145
II.1.1.2.2. The tension between the scientific and the political 149
II.1.1.2.2.1. A criterion of utility for multidisciplinary selection 149
II.1.1.2.2.2. Scientificity and activism: an oscillation between constraints 151
II.1.1.2.2.3. Epistemological issues associated to multidisciplinarity 154
II.1.1.2.3. Further risks of instrumental multidisciplinarity: misuse of conceptual tools and over-interpretation 159
II.1.1.2.3.1. The case of van Dijk's "political implicatures" 160
II.1.1.2.3.2. Passivisation as a typical CDA case 172
II.1.1.2.3.3. Presupposition in CDA 175
II.1.1.2.4. Discussion 178
II.1.1.3. Concepts and claims in mainstream CDA 179
II.1.1.3.1. Language, meaning and ideology 179
II.1.1.3.2. Further elaborations 184
II.1.1.3.3. Discourse as a social practice: issues of causality 190
II.1.1.3.3.1. Causality in CDA 192
II.1.1.3.3.2. Two notions of discourse 194
II.1.1.4. Conclusion 195
II.1.2. Argumentation theory 197
II.1.2.1. Why argumentation studies are relevant to an account of manipulative and uncooperative communication 197

II.1.2.2. Formal logic and argumentation ... 200
II.1.2.2.1. Syllogisms and formal fallacies ... 200
II.1.2.2.2. Limitations of formal logic for an account of argumentation and manipulation: Issues of meaning ... 203
II.1.2.3. Informal logic ... 207
II.1.2.3.1. Precursors: Toulmin and Perelman & Olbrechts-Tyteca ... 207
II.1.2.3.2. Fallacy theory ... 211
II.1.2.3.3. The dialectical shift ... 214
II.1.2.3.3.1. Formal dialectics ... 214
II.1.2.3.3.2. Pragma-Dialectics ... 215
II.1.2.3.3.2.1. Theoretical and practical aspects of Pragma-Dialectics ... 215
II.1.2.3.3.2.1.1 The standard pragma-dialectical model ... 215
II.1.2.3.3.2.1.2. Strategic manoeuvring as the incorporation of rhetoric in pragma-dialectics ... 219
II.1.2.3.3.2.2. PD: discussion ... 221
II.1.2.3.3.2.2.1. Two normative systems to judge soundness and fallaciousness ... 221
II.1.2.3.3.2.2.2. Pragmatics in PD ... 225
II.1.2.4. Limitations of formal and informal logic for a naturalistic account of fallaciousness ... 230
II.1.2.4.1. Objections to formal logic ... 230
II.1.2.4.2. Cognitive issues ... 235
II.1.2.5. Conclusion ... 241

II.2. CONSTRAINTS ON BEHAVIOUR ... 244
II.2.0. Introduction: influence in social psychology – its link with the study of manipulation ... 244
II.2.1. Sociologically-oriented social psychology ... 245
II.2.1.1. Milgram's experiment ... 245
II.2.1.2. Asch's experiment on conformity ... 249
II.2.1.3. Sociologically-oriented social psychology and manipulative communication ... 252
II.2.2. Psychologically-oriented social psychology and manipulation ... 255
II.2.2.1. Persuasion research: the Elaboration Likelihood Model ... 256
II.2.2.1.1. General overview ... 256
II.2.2.1.2. Discussion ... 259
II.2.2.2. Festinger's theory of cognitive dissonance ... 260
II.2.2.2.1. The theory ... 260
II.2.2.2.2. Discussion: manipulation and dissonance ... 263
II.2.3. Conclusion ... 267

II.3. COGNITIVE CONSTRAINTS ON INFORMATION-PROCESSING ... 269
II.3.0. Introduction ... 269
II.3.1. Reasoning, biases and heuristics ... 270
II.3.1.1. The Wason Selection Task Paradigm ... 271
II.3.1.1.1. Overview and contemporary reinterpretations ... 271
II.3.1.1.2. Discussion ... 277
II.3.1.2. Kahneman and Tversky's heuristic and biases programme ... 279
II.3.1.2.1. Overview ... 279
II.3.1.2.2. Discussion ... 280
II.3.1.3. Biases induced through communication ... 281
II.3.1.3.1. Concept retrieval: the Moses illusion ... 281
II.3.1.3.2. The impact of words on memory: misinformation effect and labelling effect ... 284
II.3.1.3.3. Epistemic biases: validity effect and mere exposure effect ... 286
II.3.1.3.4. Conclusion: biases, heuristics and manipulative communication ... 289
II.3.2. Filtering information: the case for epistemic vigilance, cognitive and emotional filters ... 289
II.3.2.1. Epistemic vigilance ... 289
II.3.2.2. Cognitive and emotional filtering ... 292
II.3.3. RT as a suitable framework for a cognitive account of manipulative communication ... 293

II.3.3.1. RT meets cognitive requisites 293
II.3.3.2. RT can explain pragmatic-processing-related manipulation 295

CHAPTER III. CONTEXTUAL SELECTION CONSTRAINT: TOWARDS A COGNITIVE PRAGMATIC ACCOUNT OF MANIPULATIVE, FALLACIOUS AND UNCOOPERATIVE COMMUNICATION 301

III.0. INTRODUCTION: MANIPULATION AND COGNITIVE ASPECTS OF INFORMATION-PROCESSING 302

III.1. MAIN CHALLENGES FOR A PRAGMATIC ACCOUNT OF MANIPULATION AND NON-COOPERATION 307
III.1.1. Manipulation and Speech act theory 308
III.1.2. Manipulation and Gricean frameworks 310
III.1.3. Manipulation and Relevance Theory 314
III.1.3.1. Ostensive-inferential communication 314
III.1.3.2. RT's explanatory scope: the cognitive dimension of information-processing 316
III.1.3.3. Manipulation and RT 317
III.1.3.4. Multidisciplinary advantages of a cognitive pragmatic account of manipulation 319

III.2. COMPREHENSION AS THE OUTCOME OF A FALLIBLE PROCESS 321
III.2.1. Relevance does not guarantee comprehension 322
III.2.1.1. Relevance, processing effort and cognitive effect 322
III.2.1.2. Context 324
III.2.2. Fallibility of the comprehension procedure 327

III.3. THE CONTEXTUAL SELECTION CONSTRAINT (CSC) 330
III.3.1. General overview 330
III.3.2. Salience and relevance 334

III.4. OPERATIONAL ASPECTS OF THE CSC 337
III.4.1. Strengthening strategies: increasing the salience of context C 338
III.4.1.1. Fostering accessibility of assumptions in context C 338
III.4.1.1.1. Repetition implements the CSC over time 338
III.4.1.1.1.1. The 'Stauffer Case' 339
III.4.1.1.1.2. Discussion 344
III.4.1.1.2. Emotions, processing effort and the CSC 345
III.4.1.1.2.1. Emotional refractory state and contextual constraining 345
III.4.1.1.2.1.1. The 'Sarkozy case' 347
III.4.1.1.2.1.2. Discussion 348
III.4.1.1.2.2. Argumentative appeals to emotion and contextual constraint 349
III.4.1.1.2.3. Flashbulb contexts: how 9/11 can be exploited by the CSC 351
III.4.1.2. Fostering strength of assumptions in context C 354
III.4.1.2.1. The argumentum ad verecundiam 355
III.4.1.2.1.1. The 'UDC anti-naturalisation case' 357
III.4.1.2.1.2. Discussion 361
III.4.1.2.2. The argumentum ad populum 362
III.4.2. Weakening strategies: lowering the salience of context C' 365
III.4.2.1. Weakening the salience of C' by making alternative goals salient 366
III.4.2.1.1. The 'Friday the 13th case' 367
III.4.2.1.2. Discussion 369
III.4.2.2. The argumentum ad hominem 371
III.4.2.2.1. The 'Trudeau case' 372
III.4.2.2.2. Discussion 373
III.4.2.2.3. The 'Feurer case' 375

III.4.3. Theoretical implications of the CSC model for the study of manipulation *377*
III.4.3.1. Manipulating contextual relevance and relevance to an individual: implications for IC and PC 378
III.4.3.2. Explanatory and descriptive adequacy of the CSC model 380

III.5. CSC AND ARGUMENTATION 383
III.5.1. Fallacies are not necessarily manipulative *383*
III.5.2. CSC, sound and fallacious arguments *384*
III.5.3. Concluding remarks *388*

III.6. CSC AND CDA 391

CONCLUSION 395

REFERENCES 399

INTRODUCTION

Successful communication is neither scarce nor random; people are usually very good at understanding each other, so much so that the success of actual communicative processes can be taken as a good indication that there are underlying principles governing conversational exchanges. To the extent that we can consider this tendency to be a forthright reflection of some kind of joint undertaking geared towards the fulfilment of a shared goal (namely, achieving a mutual understanding of whatever is being talked about), this is sufficient reason to assume conversational exchanges to be cooperative endeavours – to some extent at least.

Nonetheless, besides obvious cases such as refusals to answer questions or more generally abandoning the conversational exchange altogether, there are two cases of 'unsuccessful' or at least problematic communication in which the observance of these principles is compromised: i) misunderstandings, which display some kind of conventional or cognitive dysfunction, and ii) manipulative exchanges, usually characterised by a mismatch between the intentions the speaker manifests and the intentions she really has in making her utterance, typically intentions whose fulfilment require that the addressee knows nothing about them. In the first case communication is jeopardised independently of volition. In the second, the scenario is more complex: the manipulated addressee considers that communication was successful and unproblematic, whereas the manipulative speaker knows that the communicative exchange would be regarded as problematic by the addressee were the latter to become aware of the speaker's manipulative intention.

Many paradigms in the social sciences and the humanities have addressed, either directly or indirectly, the issue of manipulation and deception (see Chapter II). However, few approaches tackle the issue directly in the current scientific context: discourse-analytical traditions tend to focus on the socio/political implications of particular linguistic forms (such is the case of Critical Discourse Analysis, see §II.1.1.); argumentation theories are more interested in accounting for the acceptability of arguments, even if their treatment of fallacious argumentation is of great interest to the study of manipulation (see §II.1.2. and §III.5.); social psychology treats manipulation from the angle of behavioural influence (see §II.2.) and in this respect does not typically investigate linguistic means of achieving influence, just like cognitive psychology, even if the latter constitutes perhaps the more relevant field of study for an account of manipulative communication since it deals with the way information is processed by our brains (see §II.3.). Some references belonging to the type of cognitive pragmatic approach I will be advocating are available (e.g. Saussure 2005, Allott 2005, Blass 2005), but, as they haven't been taken on in a

consistent research agenda, they can be regarded as programmatic. To this day, no genuinely extensive cognitive pragmatic account is available. The underlying motivation and originality of this thesis is to be found in an explicit attempt to provide such an account.

Manipulative communication is the sort of phenomenon everyone has an intuitive idea about; yet, providing a list of definitional necessary and sufficient conditions is no easy task, perhaps because its identification oftentimes involves moral and ethical judgements but also speculation about intentionality. For instance, would we consider a parent who tells her/his children about Santa Claus to be manipulative (see §I.3.1.1.2. and Saussure & Schulz 2005) or in any case less manipulative than a cult guru who has managed to fill his bank account with the money unsuspecting followers have been led to give him? The identification of manipulative communication, as I shall discuss it, is a thorny question; independently of the criteria we come up with to characterise it, we will invariably find counter-examples invalidating them (see §I.3.1.). The 'what' question, as I will be led to name it further along, is therefore perhaps the most troublesome to answer. Yet I will contend that this is perhaps not the only way to account for manipulative communication.

We can also conceive of manipulative communication as a dynamic process that achieves specific effects, among which that of remaining concealed (I would consider overt manipulation to be self-defeating, and thus the expression to be a contradiction in terms) by managing to get the target of the manipulative attempt to process only a limited set of information. In other words, manipulative communication is about imposing some sort of constraint on the addressee, and in particular a constraint on the way the latter will process the information contained in the communicative stimulus. The idea of constraint in turn opens up countless directions of research. In this thesis I opt for a cognitive exploration of the underpinnings of manipulative communication and its implications on the level of human interaction. Underlying my research is therefore the idea that the success of manipulation, in communication, can be explained by taking a look at the way the addressee is (mis)led into processing the verbal stimulus uttered by the manipulative speaker in view of achieving a deceptive goal. More specifically, this thesis is about how (standard) information-processing mechanisms can be exploited to further manipulative goals.

The thesis is structured as follows: Chapter I is meant to provide a theoretical overview of the object of study, namely uncooperative and manipulative communication, so that terminological requirements of clarity and conciseness are met. Chapter II surveys available literature in order to identify relevant parameters of analysis. Chapter III synthetises these and presents an

integrated cognitive pragmatic account of manipulative communication grounded on Relevance Theory. Below is a more detailed description of the contents of each chapter.

Chapter I is a theoretical discussion of the notions of cooperation, non-cooperation and manipulation in view of grounding their study in a pragmatic framework. In order to get a full grasp of what it could mean to be manipulative in communicative exchanges – which is not the standard case of communication – a first requirement is to have an informed picture of general communicative standards, so that whether manipulative communication violates fundamental conversational principles or not can be properly evaluated (§I.1.). I take as a point of departure Grice's rationale for describing and explaining communication as a cooperative endeavour between conversational participants by assessing the implications of his *Cooperative Principle* (Grice 1989 [1975]) which essentially enjoins conversational participants to cooperate towards the success of conversational exchanges and which supposes the enforcement or exploitation of four *conversational maxims* dictating (though originally not in a normative sense) what conversational contributions should look like. I then review some pieces of research advocating a dichotomy of the Cooperative Principle on the grounds that the goals towards which conversational participants may cooperate can be of different natures (e.g. Chilton 1987, Attardo 1997a, 1997b, Gu 1993, 2003, Lumsden 2008, to name a few). I further assess this claim by drawing on recent literature in evolutionary psychology and anthropology in order to propose a further distinction between what I call *Communicative Cooperation*, *Informative Cooperation* and *Perlocutionary Cooperation*; I take the first two to be necessary conditions of communication insofar as they secure that contents are successfully exchanged, and assume the latter to exceed the scope of communication inasmuch as it concerns extra-linguistic goals that people pursue by resorting to communication. I conclude the section by noting that my clarification of the Gricean framework on cooperation allows a finer-grained description of communication, yet also acknowledge that the cooperative principle is only a descriptive convenience, as I take – following contemporary cognitive pragmatics (e.g. Sperber & Wilson 1995) – the explanation of successful communication to rest on alternative principles.

Against the backdrop of my trichotomy of the Cooperative Principle, I then examine what exactly it could have meant, in Gricean terms, to be uncooperative in a given communicative exchange by looking in detail at different ways of violating conversational maxims (§I.2.). As it turns out, both overt and covert maxim violation can be deemed cooperative in the sense that exchanges of meaning are still secured, meaning that both communicative and informative cooperation obtain whenever communication occurs – assuming that conversational participants understand each other. In other words, we cannot equate uncooperativeness (in the sense I am interested in, i.e., the manipulative and deceptive sense of non-cooperation) with maxim violation. The case of

perlocutionary cooperation is itself more elusive, because this type of cooperation is almost exclusively dependent on the conversational participants' mental states (beliefs, desires, goals, etc.) and not on operational aspects of communication. What emerges from my discussion is that, in the end, the crucial parameter characterising manipulative communication is whether the attempt is covert. Within a Gricean framework, this means that deception would be linked to the covert violation of conversational maxims. I assess this claim – which constitutes the fundamental assumption of Information Manipulation Theory (McCornack 1992) – and conclude that deception cannot be satisfactorily accounted for in terms of covert violations of conversational maxims; deception can certainly take advantage of these, but they are not necessary for manipulation to be successful.

Section I.3. then focuses on manipulation and on what it means exactly to be manipulative in a communicative exchange. I first review some definitions that have been given in existing literature before turning to elaborating my own working definitions of the phenomenon. I also try to delineate the limits between manipulation and neighbouring phenomena such as lies, persuasion and argumentation.

Chapter II is a discussion of state-of-the-art research relevant to an account of manipulative communication. From the definitions given in Chapter I (§I.3.2.), and following Saussure (2005), emerges the idea that manipulation can be envisaged as a constraint on information-processing mechanisms. If we want to assess communicative constraints, we need to look in at least three different directions: constraints that are internal to the message (i.e., constraints imposed by its form and content), constraints that are external to the message (i.e., social constraints functioning as parameters of influence) and constraints that are specific to our cognitive machinery (i.e., cognitive constraints of information-processing devices). Chapter II is thus thematically structured so as to provide a critical review of the type of work carried out in these respective directions.

The first two sub-sections of Chapter II (§II.1.1. and §II.1.2.) deal with discursive constraints and are thus devoted to Critical Discourse Analysis and Argumentation Theory respectively. Critical Discourse Analysis (CDA) is probably the most influential paradigm in the study of the type of discourse I am concerned with in this thesis, namely manipulative discourse, because it is underlain by the assumption that language is a vehicle of ideology that is instrumental to fulfilling social domination purposes. I will provide a critical assessment of CDA and be led to conclude that its methods are not suitable for the type of research I have in mind here. However, I will welcome CDA's commitment to critique and to analysis of naturally-occurring data of social and political importance as a valuable heuristic resource in the study of manipulation. I will then

review important developments of Argumentation Theory in the 20th century and assess the contribution that the type of research carried out in mainstream approaches such as the Pragma-dialectical model (cf. e.g. van Eemeren & Grootendorst 2004) can offer to a pragmatic endeavour. While one of the main goals of mainstream Argumentation Theory is compatible with part of what I am looking for, namely an account of fallacious argumentation (which can sometimes be put on a par with manipulative argumentative moves), it usually does not integrate a cognitive component in the formulation of its explanations. Both paradigms, however, provide valuable resources to understand the type of constraints that messages may carry.

The second section of Chapter II (§II.2) deals with constraints on behaviour accounted for within Social Psychology. Mention will be made to famous experiments on conformity by Asch (1955) and on authority by Milgram (1974). The idea is to see whether the type of influence coming from outside the individual can contribute to explaining why manipulation may work. I will argue that the principles and mechanisms uncovered in social-psychological research (either sociologically- or psychologically-oriented) have linguistic, argumentative and discursive counterparts, and that as such they provide reliable empirical evidence for the type of manipulative mechanisms that will be described in Chapter III.

The final section of Chapter II is devoted to an assessment of internal constraints and accordingly incorporates a review of literature in cognitive psychology. The idea is to report striking findings supporting the idea that our minds, unlike computers, are not systematic but fallible and imperfect information processors. We usually take cognitive shortcuts to process information, which makes us prone to falling victim of cognitive illusions. In other words, there are cognitive constraints such as heuristics and biases which determine the output of information-processing.[1] This is crucial to my account of manipulation since I hypothesise that the success of a manipulative attempt depends on whether the speaker manages to exploit the addressee's interpretative mechanisms in such a way that is compatible with the realisation of the concealed manipulative intention. A case will be made for the interpretation of manipulation in terms of information-processing, notably by drawing on recent work on credulity (Clément 2006) and epistemic vigilance (Sperber *et al*. 2010). Additionally, I will show that the relevance-

[1] I here use the term *bias* in its technical acceptation meant to denote a tendency for mental processes to make errors and therefore depart from otherwise habitually accurate and adequate information processing. I will thus deliberately avoid a discussion about the notion of bias in terms of ethical, social or formal norms. In line with cognitive psychological research on cognitive biases, I take the latter to be responsible for cognitive illusions (see Pohl 2004), which are therefore explained similarly to optical illusions.

theoretic framework is well-suited to integrate the findings of these disciplines into an interfaced cognitive pragmatic account of manipulative communication.

Chapter III starts with a discussion of different options for a pragmatic study of manipulation and focuses on the challenge faced by any pragmatic account, namely that a phenomenon whose success requires that its motivating intention remain concealed cannot, in principle, be accounted for within a theory which holds intention-recovery to be the pivot of successful communication. I conclude this first sub-section by highlighting some aspects of Relevance Theory that in my view will overcome this difficulty and at the same time provide decisive arguments in favour of an RT-based account of manipulative communication. The rest of the chapter, in this respect, represents my contribution to the study of manipulation and presents the details of the Contextual Selection Constraint model (cf. Maillat & Oswald 2009, forth.). Underlying the model is the idea that manipulation is a twofold process which i) forces the hearer to interpret the message within a limited context and ii) prevents the hearer from accessing other critical contextual information that would be highly detrimental to the manipulative attempt altogether. Similarly to magicians/illusionists/pickpockets, who manage to focus your attention on one hand while they trick/rob you with the other, manipulators focus your attention of certain information so that other critical information is left unprocessed. The model heavily builds on Relevance Theory's conceptions of relevance and context and assumes that salience of information may be constrained, either by strengthening strategies (ensuring that only accessible/strong information is processed) or by weakening strategies (meant to lower the salience of critical information that would defeat the manipulative attempt). For illustrative purposes, I provide several documented examples of these strategies to describe and explain how they achieve their effect. The last sections of Chapter III (§.III.5. and §III.6.) constitute a broader opening towards Argumentation Theory and CDA. They point to a theoretical convergence between the proposed model and some mainstream conceptions about argumentation and fallacies widely shared by researchers within the field of Argumentation Theory, but also argue in favour of a more cognitive CDA.

The theoretical and methodological underpinnings of this research require a hypothetical-deductive epistemology of a mechanistic and naturalistic kind which is characteristic of research in cognitive science, whereby the issues that will be addressed are meant to reflect how the object of study behaves naturally and spontaneously in the world. Moreover, I will follow a model-theoretic approach stipulating that being able to provide a model of cognitive processes, more specifically a model of how these processes are mechanistically articulated in terms of causal relationships, amounts to explaining those processes. In particular, I will be led to analyse processes of comprehension which take as input verbal stimuli produced in communicative

exchanges and transform it into mental representations about the world. Manipulative communication, in this sense, will be modelled accordingly so as to specify how the supply of specific inputs in specific contexts might lead to the formation of mental representations that should normally be rejected under critical scrutiny.

As a consequence, I will adopt the perspective of the hearer, assuming that a fruitful method to account for manipulation is to reflect on what addressees go through, in terms of information-processing, when they process manipulative stimuli. In this sense, my analysis will make use of my own intuitions about language use (but I will also summon experimental research in cognitive psychology to ground these intuitions in experimental data). The corpus I selected for analysis is either made of imaginary examples taken for illustrative purposes and licensed by linguistic intuitions or of attested examples whose uncooperative or manipulative nature has been independently corroborated by other researchers.

Underlying the model is also the assumption that humans are equipped with a *theory of mind*, also referred to as *folk psychology* or *naïve psychology*; namely, humans are naturally able to attribute mental states (beliefs, desires, intentions) to others by way of metarepresentation (see Sperber 2000a). As noted by Clément (2006), the phenomenon of manipulation is perhaps the most vivid illustration of this metarepresentational ability and of the existence of the theory of mind: only when a manipulator is able to anticipate the effect her words will have on her addressee's mental states will she have some prospect of being successful in her manipulative attempt. The strategies she will deploy to attain her covert goal are presumably not random: they are chosen precisely because they are deemed to affect the target's mental states, and this kind of design necessitates a theory of mind. Furthermore, if we choose to focus on how information is processed under manipulative conditions, then we might also, via the fundamental theory of mind assumption, end up indirectly assessing the manipulative intention on the speaker's behalf. Focusing the model on the effects manipulation can have in terms of information-processing is therefore one possible way of assessing a speaker's model of intentional manipulative communication.

The general principles I will use to describe uncooperative and manipulative communication correspond to some founding principles of cognitive science and of Relevance Theory in particular: principle of relevance, principle of non-demonstrative rational deduction, principle of risk in processing information, principle of intention recognition. Accordingly, I will work with the crucial assumption that manipulative communication is an intentional phenomenon, which allows me to consider it as part of a cause-to-effect relationship, and hence as a possible object of study from a cognitive perspective. Should this not be the case – i.e., should we admit that

manipulation is some kind of unintentional effect of an act of communication –, we would be faced with an impossible postulate escaping the laws of causality by virtue of its fortuitous and unpredictable nature. In other words, should we reduce the manipulative effect to an accidental by-product of communication, its manifestation would be random and any identification of its conditions of realisation would be impossible. In what follows I take on the opposite view and provide the reasons for which I regard manipulative communication to be accountable for within a cognitive pragmatic programme.

CHAPTER I. COOPERATION, NON-COOPERATION AND MANIPULATION: THEORETICAL OVERVIEW

"The only thing that will redeem mankind is cooperation". Bertrand Russell

I.0. INTRODUCTION

The primary aim of this chapter is to provide a panorama of the notion of cooperation in the field of Pragmatics, and in particular in H.P. Grice's framework, who was the first to postulate cooperation as a definitional feature of human communication. Such a panorama is needed in order to start considering fundamental issues of non-cooperation, manipulation and deception in human communication.

The first section (§I.1., "Cooperation") of this chapter will thus be devoted to an exposition of the fundamental tenets of the Gricean model; I will review some of the post-Gricean literature which has refined the so-called Cooperative Principle and propose that the latter should be broken down, for descriptive purposes, into three levels of cooperation. In doing so, I will be led to examine ideas stemming from research in evolutionary psychology and anthropology as well as work within game theory in order to elaborate an informed picture of the notion of cooperation in an account of communication.

The second section (§I.2., "Non-cooperation") will put the findings of the first section to use by considering what it could have meant, for Grice, not to cooperate in conversational exchanges. The overarching goal of this thesis being to supply an account of deceptive and manipulative communication, I will accordingly try to critically assess the usefulness and adequacy of Gricean Pragmatics for such an endeavour.

The third section of this chapter (§I.3., "Manipulation) will on the one hand provide a theoretical discussion about the notions of manipulation, deception and non-cooperation so as to elaborate working definitions of each, and on the other hand will assess the relationship of manipulation with neighbouring phenomena such as argumentation and persuasion, in order to conclude the theoretical discussion of my object of study.

I.1. COOPERATION

I.1.0. OVERVIEW

This section examines the main claims that have been made in pragmatic literature on cooperative communication since the seminal work of Herbert Paul Grice in the philosophy of language. Starting from Grice's original account of the Cooperative Principle (henceforth CP), I review some refinements that have been adduced to his theory, notably by critically assessing the dichotomy researchers have operated on the CP. Building on work in evolutionary psychology, anthropology and game theory, I then argue that in fact, it makes sense to further break down the widespread dichotomy of the CP into a tripartite conception of cooperation – in terms of communicative cooperation that is – so as to incorporate a very rudimentary and basic

level of cooperation in the account, as a natural and rational disposition of individuals to communicate. The aim of these developments is to characterise the notion of cooperation proper, and to describe what it means exactly to be cooperative in communicative exchanges.

Following recent work within the framework of Relevance Theory (henceforth RT), I will then suggest that a theory of communication benefits from such a tripartite conception only as far as description is concerned; in what regards the explanatory level of the theory, however, I will build on RT's criticism of the CP to argue that such a principle does not satisfactorily account for the way specific meanings are conveyed in communicative exchanges. This will lead me to support the descriptive value of cooperation, but to reject it as a sufficient explanatory principle.

In so doing, I will set the grounds for a discussion on non-cooperation, manipulation and deception, to be found in the subsequent sub-chapters.

I.1.1. GRICE'S RATIONALE FOR COOPERATION

I.1.1.1. General remarks

The language-related acceptation of cooperation, which is nowadays incorporated in many pragmatic theories as denoting a default feature of communicative interaction, traces back to the seminal works of H.P. Grice.[2] Although much of the contemporary pragmatic literature heavily builds on his Cooperative Principle (CP) and his maxims of conversation, Grice's original focus was not exclusively on cooperation proper. In fact, Davies (2000: 1) argues that the concept of cooperation is "peripheral to [Grice's] thought", and judging by the introductory pages of his (1975) paper ("Logic and conversation"), it indeed seems that Grice's primary ambition was to overcome the limitations formal logic encounters when faced with issues of meaning and comprehension in natural language, but not to provide a pragmatic account of cooperation proper. We could add that the way the CP is originally phrased by Grice (1989: 26) as well as the layout employed to expose it in the article do not suggest that Grice himself thought of the CP as a paramount principle of communication; in fact, he meant it to account for implicature derivation.[3]

From a wider perspective, and this is Davies' (2000) take on the matter, his work is actually concerned with communicative *rationality*; just as logicians elaborated a formal system for the logic of propositions, Grice thought there was a logic of conversations which followed certain rational standards that humans observe – or exploit for that matter – when they communicate.

[2] See Grice's (1989 [1975]) "Logic and conversation".

[3] See Davies (2000) for this specific take on the role of cooperation in Grice's work.

Grice's philosophy is thus not so much about cooperation, but about rationality. To quote Davies again: "Grice sees cooperation as the necessary outcome from the application of reason to the process of talk. In other words, cooperation is the realisation of rationality applied at the level of discourse" (Davies 2000: 19).

However, this is not to say that cooperation is only secondary in Grice's philosophy; it is central in his account of implicature, since, according to his model, it is by virtue of the assumptions hearers may make about speaker cooperativeness that implicit meanings are successfully communicated.

Grice's view is that communication among humans is purposeful and rationally driven. He believes that rational communicative conduct is manifested through some kind of cooperation between language users, and accordingly postulates the existence of a Cooperative Principle which he formulates as follows:

(1) Grice's Cooperative Principle:

> "Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk in which you are engaged." (Grice 1989: 26)

Grice also highlights the cooperative nature of communication by considering "cooperative efforts" to be something language users are willing to make,[4] as "each participant recognizes in them, to some extent, a common purpose or set of purposes, or at least a mutually accepted direction" (Grice 1989: 26) in the conversation. Grice's account holds that such a rational behaviour is expressed by the use (i.e., either observance or overt violation) of four conversational maxims which yield results in accordance to the CP. These are Quantity, Quality, Relation and Manner (which comprises four sub-maxims), and are expressed as follows:

(2) Grice's conversational maxims:

> "Quantity:
>
> Make your contribution as informative as is required (for the current purposes of the exchange).
>
> Do not make your contribution more informative than is required.
>
> Quality: Try to make your contribution one that is true.
>
> Do not say what you believe to be false.

[4] It is usually assumed that these processes are unconscious and non-reflexive.

Do not say that for which you lack adequate evidence.

Relation: Be relevant.

Manner: Be perspicuous.

1. Avoid obscurity of expression.

2. Avoid ambiguity.

3. Be brief (avoid unnecessary prolixity).

4. Be orderly." (Grice 1989: 26-27)

The CP stipulates that the speaker rationally makes use of these maxims; accordingly, the hearer will infer speaker meaning based on these and on the assumption that the speaker is complying with them. The account Grice proposes ultimately aims at providing a theoretical account of meaning which is compatible with (i) the assumption that conversations are not random sequences of utterances and that verbal exchanges do follow certain standards, and (ii) the idea that speaker meaning is arrived at not through mere decoding but through inference.

A great deal of research in pragmatic literature stemming from Gricean theory has focused on these rational standards of communication (i.e., the conversational maxims), generating roughly two major orientations in subsequent pragmatic work: neo-Griceans, such as Levinson (1983, 2000), Horn (1984), Bach & Harnish (1979), have pursued Grice's line of work in an effort to specify the maxims and account for the principles and standards language users are deemed to observe in ordinary conversation. Post-Griceans on the other hand, the major proponents of which are Relevance theorists (following Sperber & Wilson 1986, 1995), took on the inferential nature of communication defended by Grice in a reductionist programme that advocates, on cognitive grounds, that communication is guided by a unique principle, the principle of relevance. This account was thus designed to reduce the complexity of the Gricean model, and argued against both the theoretical usefulness of four maxims and the cognitive implausibility of the Gricean 'working out schema' according to which meaning is arrived at.[5]

However, explicit and exclusive focus on cooperation proper has not been the main concern of pragmaticians ever since, to the extent that very few of them actually attempted to design a formal theory of cooperation in communication.[6] Post-Griceans on the one hand suggest that this is because we do not need a CP in order to account for communication. Neo-Griceans on the

[5] See Sperber & Wilson (1995: 161-162) and Wilson & Sperber (1981) for a critique of Grice's account.

[6] Although we do find exceptions, such as Attardo (1997a, 1997b) and Gu (1993, 2003), who have addressed more in detail the cooperative principle and its implications for communication.

other hand treat cooperation as a default principle language users are deemed to comply with in their verbal exchanges, but seldom engage in a thorough discussion of the CP's nature and implications for an account of communication, somehow taking cooperation for granted; they usually turn their attention to conversational maxims and the issue of implicature.

Nonetheless, among the few authors who did specifically address cooperation, some sort of consensus has been achieved as to the nature and kinds of cooperative behaviour one may be willing to engage in as one communicates; notably, we will see that a dichotomy of Grice's CP is fairly accepted today (see §I.1.2.).

Outside pragmatics, research on cooperation in the anthropological sense – i.e., taken as a behavioural option in the social organisation of hominids – has developed for many years in the fields of evolutionary psychology, game theory, artificial intelligence and anthropology. Most of these paradigms defend the idea that cooperation is a rational (to a certain extent at least) and advantageous behaviour for human beings. We will survey some of this research on cooperation further along (see §I.1.5.), in order to assess the relationship between what can be considered as rational behaviour and its actual repercussions for communication, but let us for the moment focus on the notion of cooperation proper, as formulated by Grice.

As noted by Attardo (1997a, 1997b), Chilton (1987) and Lumsden (2008) among others, Grice's formulation of the CP raises a problem of ambiguity in that it can be taken to conflate two levels of cooperation, to which we shall return shortly. Roughly speaking, we can identify a communicative level of cooperation, which is centred on the exchange of meaning, and an extra-communicative level of cooperation, which is determined by the goals people can pursue and tentatively attain *as a result* of successful communication. In his work, Grice barely evokes the two facets of the issue, although the examples he discusses clearly point to it.

In all fairness, however, it must also be acknowledged that Grice was aware of the possibility of broadening the notion of cooperation as the following quote from "Logic and Conversation" (Grice [1975] 1989) illustrates:

> "I have stated my maxims as if this purpose were a maximally effective exchange of information; this specification is, of course, too narrow, and the scheme needs to be generalized to allow for such general purposes as influencing or directing the actions of others." (Grice 1989: 28)

In Strand Six of his "Retrospective Epilogue", Grice also acknowledges that his Cooperative Principle must not be taken in such an exclusive way:

> "So since we are concerned as theorists only with concerted talking, we should recognize that within the dimension of voluntary exchanges (which are all that concern us)

> collaboration in achieving exchange of information or the institution of decisions may coexist with a high degree of reserve, hostility and chicanery and with a high degree of diversity in the motivations underlying quite meagre common objectives. […] but such exchanges honour the cooperative principle at least to the extent of aping its application." (Grice 1989: 369-370)

We will see further along that research in pragmatics has taken this direction seriously and has developed an account of cooperation which does justice to the various ways cooperation can be conceived of (see in particular §I.1.4.).

Conversational maxims, together with the CP, are deemed by Grice to serve purposes of various kinds, from a purely communicative purpose (the "maximally effective exchange of information") to a more social purpose, such as, say, influencing someone else's behaviour through communicative exchanges. The remainder of this section will discuss the implications of this possibility for an account of cooperative communicative behaviour.

I.1.1.2. Implications of cooperation

Common-sense (and dictionaries) tells us that to *cooperate* is to work jointly towards the same end/goal. Cooperation therefore requires at least two participants as well as some kind of shared goal that both are willing to attain. As Attardo (2003: 4) notes, "goal adoption is of course cooperation". Furthermore, the definition entails that a goal is attainable only to the extent that both participants i) share this goal (in the sense that both are willing to attain it) and ii) make efforts – or engage in a relevant course of action – for it to obtain. This also builds on the idea that the attainment of any goal supposes the adoption of a purposeful behaviour. The consequence of cooperative behaviour is arguably some kind of benefit that both participants can be taken to earn from cooperative action; and in fact, it may very well be the case that this prospective benefit serves as a motivation – or purpose – to engage in cooperative behaviour in the first place. It is important to keep in mind that cooperation supposes a shared goal between participants and that it carries with it a sort of guarantee that some benefit will be gained through this particular course of action (together with the idea that this benefit would not be obtained, or obtained at a greater expense, in the absence of cooperative action), and thus that such an endeavour requires that participants make efforts. Accordingly, I would not be opposed to considering that a cost/benefit dynamics regulates cooperation, similarly to the sort of dynamics we find for instance in Relevance Theory – albeit without reference to cooperation in that particular framework –; this in turn might constitute an interesting starting point for the formulation of specific assumptions about the role of cooperation in communication, and, perhaps, with regard to the cognitive processes involved in communication.

Accordingly, I will explore two distinct aspects of cooperation that could in my opinion contribute to the description of communicative exchanges: the first is the scope of cooperative behaviour in terms of the dimensions of communication it concerns, while the second relates to the nature of the goals one may pursue through cooperative behaviour (in terms of communication). My own goal, ultimately, is to promote the idea that there are grounds to advocate a tripartite conception of cooperation which takes into account at the same time the purposes of communication and the communicative requirements the fulfilment of such purposes calls for.

I.1.2. COOPERATION FOR COMMUNICATION SIMPLICITER: COMMUNICATIVE COOPERATION

I.1.2.1. The cooperative paradox and the dichotomy of the CP

In his (1987) paper, Paul Chilton addresses the issue of what he terms "the cooperative paradox", which expresses the somewhat puzzling, yet intuitively conceivable, idea that inferring that a speaker is being uncooperative itself requires recourse to a CP. To take an example, one can infer from a speaker's contribution that she is being evasive, for instance by not providing a relevant answer to a question (i.e. by violating the relation maxim), only if one is able to *interpret* whatever utterance she produced, as in (3) (inspired from Lumsden 2008: 1903):

(3) Journalist: "Do you think the recent economic crisis will force the Secretary to resign?"

Politician: "No comment"

The politician is being uncooperative in her reply by deliberately trying to avoid fulfilling the journalist's goal, which is to collect information about her opinion on the content of the question (i.e., whether the politician thinks that the Secretary will resign). Yet, one may infer that the politician is being uncooperative only to the extent that one is able i) to understand what "no comment" means, and ii) to enrich that content so as to yield a representation such as for instance "the speaker does not wish to/cannot answer this question", which turns out to be uncooperative only with respect to the journalist's initial goal of knowing what the politician thinks will happen to the Secretary. The point is that even if uncooperativeness obtains, in terms of the outcome of this particular exchange, such an interpretation can only be inferred by first understanding the politician's utterance and its implications. If one is not able to understand anything in the first place, there is little chance one might infer anything at all about any of the speaker's intentions, let alone about her degree of cooperativeness in the exchange. The only thing one would be able to infer is that the speaker intended to communicate something, whatever that may be. But this is not sufficient for communication to be successful.

There is room to discuss whether the politician's utterance in (3) is actually an uncooperative answer or not; one probable inference a hearer may draw from the answer is that she actually knows what will happen to the Secretary, but that professional obligations forbid her to say it. Moreover, whether the answer was accompanied by characteristic prosody, gestures and facial expressions might trigger further inferences on behalf of the addressee, who could very well be led to interpret the answer as a yes or a no. However, if we take the answer at face-value, it communicates that the speaker chooses not to cooperate towards the goal of making what she knows public, which was the point of the question the journalist asked her. With respect to this goal, thus, and unless contextual cues prompt for further inferences as evoked above, the answer is uncooperative as it is not instrumental to fulfil the journalist's goal.

Hurford (2007: 252-255) discusses deliberate deception to argue that even if communicative exchanges turn out to be overall uncooperative, they are still cooperative on at least one level, namely the level of meaning:

> "There is still a crucial sense in which cooperation between speaker and hearer is involved in all truly communicative exchanges. Even in an act of deliberate deception, the deceiving speaker at least cooperates to the extent that he couches his message in a code which the hearer understands; there is momentary cooperation in the very transmission of the message, even if the longer-term goal of the speaker is not helpful to the hearer. Momentarily, during the production and parsing of the deceiving utterance, speaker and hearer are playing the same game together." (Hurford 2007: 253)

Again, following Hurford's example, it seems that non-cooperation still supposes cooperation (albeit of a different kind); the paradox of cooperation thus involves the idea that uncooperativeness relies on cooperativeness.[7]

The way out of the paradox, following Chilton, would hence be to postulate two levels of cooperation. As he notes:

> "It is important for our purposes to distinguish between the CP viewed as a communication requirement, a *sine qua non* for the making of meaning, and the CP viewed

[7] It could be objected that not answering at all a given question is uncooperative and does not rely on communicative cooperation; however, to the extent that not answering a question is something done manifestly, i.e., with the intention of being recognised, then we could say that silence is cooperative. Manifest silence, just as manifest utterances, is cooperative (although not in a linguistic sense) because it communicates to the addressee that the speaker is deliberately refusing to communicate; this way, the message that there will be no message still goes through (unless, of course, the speaker does not answer because she did not notice that a question was addressed to her, but in this case, we are not dealing with a case of communication).

> as an ethical, cultural, social or political imperative. [...] On the one hand, we distinguish the CP as a minimal requirement for human communication of any type. On the other hand, we distinguish the CP as a requirement imposed in specific types of communication." (Chilton 1987: 222)

One way to develop Chilton's view would be to reason in terms of the different types of goals towards which language users can be led to cooperate. An utterance may indeed serve as a basis for the fulfilment of different purposes in communication: by uttering any content, one may just be trying to secure the channel of communication – plainly phatically – by getting the hearer to acknowledge that further exchanges are to take place (as when one says 'hello' when picking up the phone); one may try to get someone else to believe something by engaging in an argumentative or persuasive exchange, or to behave in a certain way which would be compatible with one's extra-linguistic goals (as is the case when people request favours from others); and so on. The types of goals speakers may pursue in communication can therefore be of different natures. It is precisely with respect to the criterion of the nature of these goals that Chilton's bipartite conception of cooperation is interesting.

Accordingly, a first, very rudimentary, minimal and basic level of cooperation could be envisaged in the CP. Chilton postulates that this restricted CP is designed to ensure communication *simpliciter*, defined as the situation where "the hearer assumes the speaker assumes the hearer… assumes that the speaker is observing the CP" (Chilton 1987: 224). Thus this is a *functional* CP which ensures that communication can take place. Put simply, *functional cooperation* ensures that meaning is transmitted successfully, thus echoing Grice's idea of a "maximally effective exchange of information" (Grice 1989: 28), independently of the consequences the exchange will have in the world, beyond the communicative exchange. As it stands, the functional CP is a definitional feature of meaningful verbal contributions and exchanges, which allows "distinguish[ing] communication from gibberish" (Chilton 1987: 225). And it is indeed quite straightforward to assume that as a condition for any communicative act, the intended meaning, as a representation, must first be mutually manifest to, and interpretable by, both participants. It is thus not a far-fetched idea to construe this minimal requirement of communication as cooperation, in so far as it is in the interest of communicative participants to be willing to attain the goal of an exchange of information (whatever that information might be). The rationale for this claim could be developed as follows:

(4) Minimal (functional) cooperation as a shared goal:

- A speaker who is willing to engage in communication can be said to have the goal of having whatever she utters understood by the hearer.

- A hearer who is willing to engage in communication can be said to have the goal of understanding whatever the speaker has uttered.

- Accordingly, each of them will make cognitive efforts in order to achieve this goal: the speaker by formulating an utterance which, under rational standards, is interpretable; the hearer by processing and interpreting the utterance, with the expectation that the speaker observed rational standards of communication in making her utterance.

- Communication, minimally construed as the transmission of meaning, can only take place if the above conditions are met.

Functional cooperation so construed is the minimal requirement for meaning to be successfully conveyed. As such, and in Austinian terms, Chilton's functional CP bears locutionary value and at the same time carries the *expectation* that the content will be illocutionarily meaningful, somehow prior to any consideration about the actual, context-dependent, illocutionary value of the utterance.

I.1.2.2. Communicative cooperation

The reference to Austinian concepts such as *locution* echoes Attardo's (1997a, 1997b, 1999) account of cooperation. He also proposes to dichotomise Grice's CP by distinguishing locutionary cooperation and perlocutionary cooperation.

Locutionary Cooperation (achieved by observance of the Locutionary Cooperative Principle), is "the amount of cooperation, based on the CP, that two speakers must put into the text in order to encode and decode its intended meaning" (Attardo 1997a: 756). Perlocutionary Cooperation (achieved by observance of the Perlocutionary Cooperative Principle), is "the amount of cooperation two speakers must put into the text/situation in order to achieve the goals that the speaker (and/or the hearer) wanted to achieve with the utterance" (*ibid*.).[8] We shall return to this second level of cooperation further along, but for the time being, let us examine in more detail locutionary cooperation.

"Understanding any sentence always involves LC [Locutionary Cooperation]", says Attardo (*ibid*.). This is consonant with the definition of functional cooperation proposed by Chilton. Both accounts indeed assume that this first level of cooperation is compulsory for communication to

[8] It should be noted that Attardo's formulation leaves room to question the measurability of cooperation. To the extent that he speaks of the "amount of cooperation" one may put into communication, we may wonder if there is a way to measure this amount.

take place. It is worth also noticing that both authors make use of the Austinian classification of speech acts: Attardo (1997a) by naming the two levels of cooperation "locutionary" and "perlocutionary", and Chilton by noting that

> "any utterance can be seen *both* as a **locution**, the simple understanding of which requires the functional *assumption* of co-operation, *and* at the same time as an act that can be judged with respect to general expectations of rational co-operation underlying all forms of action, verbal and non-verbal alike." (Chilton 1987: 226, author's italics, my bold)

It is striking that none of them uses the term *illocution* to describe the levels of cooperation they argue for. In fact, it seems that they conflate Austin's locutionary and illocutionary aspects of speech acts into their first level of cooperation. One could wonder why this is so. If we refer back to Austin's definition of locutionary act, what locution describes can be roughly defined in terms of semantic meaning and reference, along with verbal, syntactic and phonetic aspects of the utterance, which are to a large extent code-related aspects of meaning. In other words, locution originally scopes over *semantic* meaning, ensuring a somewhat minimal layer of comprehension.

Now, it is only at the level of *illocution* that one can start addressing the issue of full-fledged speaker *intended* meaning, semantic underdeterminacy, and hence matters of implicit content. The level of illocution, reinterpreted in Gricean terms, should therefore correspond to the level of operation of the conversational maxims, since it is at this stage that the calculation of implicatures takes place. However, in Attardo's and Chilton's dichotomy of cooperation discussed in the present section, it appears that illocution is subsumed into functional/locutionary cooperation. Chilton notes in that respect that "it is the functional CP (*Maxim of Relation*) that permits […] implicature" (Chilton 1987: 227, my italics), which corroborates the idea that he includes illocution in his functional CP. As for Attardo's account, the locutionary cooperative principle includes a reference to the derivation of *intended* meaning along the same lines, and therefore encompasses both locution and illocution. Probably the reason for this conflation has to do with the difficulty of establishing the relevance of locutionary aspects of an utterance (taken in isolation) in an account of communication, in that they are limited if we are to consider full-fledged speaker intended meaning, notwithstanding the fact that locution already has to do with (semantic) meaning.

Thus, for the sake of observing the original Austinian distinction between locution and illocution, one could in principle further break down functional/locutionary cooperation into *locutionary* and *illocutionary* cooperation. This restricted interpretation of Attardo's locutionary cooperation resonates with the argument presented in (4) and is closer to Austin's definition of what a locutionary act is. In Gricean terms, (4) does not require the type of cooperation referred to here to be associated with any operation of the conversational maxims; in other words, and

insofar as the Austinian acceptation of locution is taken to be limited to code-bound elements, there is no reason to incorporate into locutionary cooperation cases of communication of implicit contents, because these are intention-dependent and thus fall outside the scope of locution. In fact, there is *a priori* no way to derive conversational implicatures on the basis of a locution alone, because conversational implicatures derivation goes beyond code-bound contents.

This calls for a restricted construal of locutionary cooperation. I shall follow Hurford's terminology (see Hurford 2007: 306) here and call it *communicative cooperation* (henceforth CC). It would have been tempting to rename Attardo's *locutionary cooperation* in order to call it *illocutionary cooperation* (since it involves processes of both implicit and explicit meaning construction); this would have left the expression *locutionary cooperation* available for me to denote this first, minimal, level of cooperation. However, it would have made me use 'locution' in a sense which does not do justice to the Austinian spirit, because locution, in Austinian terms, already incorporates aspects of meaning (i.e. semantic ones). As my minimal level of CC is limited to some kind of rational readiness or disposition to communicate, prior to any matters of actual meaning, I would have had to use the term in an unnecessarily altered sense.

This first level of cooperation would thus account for the mere *possibility* that meanings may be conveyed in conversation, without any claim as to the specific processes involved in deriving full-fledged meanings. In short, this minimal level of cooperation is some kind of default assumption language users make about the 'upcoming' and expected meaningfulness of locutions (or verbal contributions) in the conversation. Recall here that Chilton construes functional cooperation as a means of distinguishing meaningful communication from pure gibberish. In this sense, CC, more than a principle which is postulated to explain how it is that we are able to understand communicative stimuli, denotes one single, rational (to an extent I will try to characterise further along, see §I.1.2.3), condition required for comprehension to take place, prior to any cognitive calculation of meaning, namely that communication requires a speaker to make an interpretable utterance and a hearer to interpret it. CC as such is a necessary, though not sufficient, condition for successful communication; it is by taking for granted that the speaker is CC-cooperative that the hearer will be led to engage in cognitive processing of the utterances which convey speaker meaning. It would indeed be pointless to try to understand a stimulus without the expectation that the latter is meaningful. One could also interpret CC to require both the *expectation of rational behaviour in communication* and the *readiness to communicate*, prior to any specific consideration about actual mechanisms of comprehension.

The conception of CC can be linked to one of the four degrees of cooperation addressed by Gu (2003). He defends the idea that in talk exchange there are four degrees of cooperation, which can be grouped in two classes: pragmatic and rhetorical cooperation. The first two degrees of pragmatic cooperation include i) an initial willingness to listen to each other, and ii) the fact that the speaker says something to the hearer while the hearer interprets it (this degree of cooperation intervenes at the 'Speaker sending / Hearer interpreting' level of interaction, which can be taken to correspond to Chilton's communication *simpliciter*). The other two degrees of rhetorical cooperation concern the actual exchange of information and the nonverbal outcome the exchange might have; as a consequence, they pertain to the category of extra-linguistic cooperation, which I will deal with further along.

Gu also seems to posit the capital importance of this first degree of cooperation as he formulates cooperation in terms of purpose and goal. His classification accounts for cooperation as a degree-sensitive notion. For instance, one can be said to cooperate in a very restricted sense (the first degree) if one merely understands a question a speaker asks, insofar as one is contributing to making sure that whatever content the speaker uttered is picked up and interpreted. Now, Gu still qualifies the second degree of cooperation as involving the hearer's *interpretation* of the message; this leaves room both for semantic and pragmatic processing on behalf of the hearer (i.e. locution and illocution, in Austinian terms). Thus, if we are to develop the notion of CC as outlined above, we would need to break down Gu's first two degrees of cooperation, in order to end up with only the restricted cooperative requirement. Again, this would be a very minimal degree of cooperation: if the hearer does not understand the speaker's question, little if no interaction at all around and about the propositional content can take place. This is precisely what Gu refers to when noting that "the interaction of S-sending/H-interpreting information *is still not a talk exchange* (Gu 2003: 47-48, my italics), which requires H to provide information in response to S's." We find here a confirmation of the idea that CC is a necessary condition for communication if we consider communication to involve at least two contributions, the second being a conversational consequence of the first.

We end up with a specification of CC as follows (this echoes the last point of (4) above and Gu's (2003) characterisation of the first degree of cooperation):

(5) Communicative Cooperation (CC):

- The speaker must be willing to communicate something to the hearer.
- The hearer must be ready and disposed to listen to the speaker and interpret her utterance.
- The speaker's utterance must be interpretable (i.e. rationally designed).

- The hearer must assume that the speaker's utterance is interpretable.

From the looks of this characterisation, it seems we are definitely dealing with a necessary condition for communication. This is in fact what I advocate: CC is a necessary condition for communication, and I will further hypothesise that it is embedded in our rational communicative apparatus. We will find a close idea in a Relevance-theoretic account such as Allott's,[9] although without recourse to the concept of cooperation.

Hurford hints to this level of cooperation, but specifies it by making it dependent on the linguistic code:

> "To be communicatively cooperative is simply to use the same communicative code as one's interlocutor. Given communicative cooperation, a speaker is free to decide whether to be materially cooperative (i.e. to give true relevant information in a clear manner) or not (i.e. to lie or obfuscate). Communicative cooperation is 'playing the same language game'." (Hurford 2007: 305-306)

His take on CC is framed in an evolutionary perspective, and aims at resolving the issue of the origin of meaning and communication within social groups. He reduces CC to conversational participants sharing and making use of a same communicative code. I will shortly elaborate on this view in order to link it with human rationality in communicating, which is the subject of the next section.

Summing up, I do not consider CC to involve actual considerations about meaning (I will argue below that informative cooperation is the kind of cooperation concerned with both explicit and implicit aspects of meaning). Accordingly, my take has more to do, quasi phatically and somehow materially (in the sense that it 'materially' makes possible the exchange of meaning), with the possibility of meaning and the psychological and biological conditions that make communication possible. I try to characterise more precisely CC in what follows.

I.1.2.3. Communicative cooperation as rationality of communication

Having defined CC, two questions nonetheless remain.

First, is this level of cooperation relevant in an account of communication, to the extent that comprehension takes place in the hearer's mind, and does not require, in order to be successful, the hearer to communicate that it has indeed taken place? In other words, does this first degree of cooperation require mutual recognition? This amounts to questioning the cooperative status

[9] See §I.1.7.2.

of comprehension, and, in a way, the relevance of CC with regard to communicative rationality, since we are dealing here with what language users can reasonably be taken to expect from each other's behaviour.

However, if we reason in terms of the goals that can be attained in and through communication, we could be led to formulate an adapted answer to the problem: in so far as the processing hearer is indeed contributing to fulfilling the speaker's goal of having her utterance understood and considering that unless *any* hearer is doing this, this particular speaker goal will never be fulfilled, we will assume, following Gu (1993, 2003) – though reducing his take to CC – that it does make sense to describe this behaviour as cooperative, albeit in a rudimentary way. In short, I legitimise the relevance of cooperation only to the extent that both speaker and hearer can be said to pursue the same goal: they both engage in a communicative exchange so as to end up entertaining a similar representation and they truly act in view of fulfilling this goal. CC thus supposes goal adoption (even if this might be totally unconscious) from conversational participants, and thus goal sharing. However, it implies that we may speak of cooperation to characterise behaviour which is not necessarily (actually, practically never) conscious or self-reflexive. This could be a reason to construe CC as an inherently rational feature of our minds, in that a readiness to communicate and to understand can both be taken to be means – what is more, efficient means – to fulfil a communicative goal.

The second problem, briefly evoked above, that arises if we take on board the idea that CC plays a role in communication is that postulating the latter may turn out to be completely trivial, because every communicative exchange, no matter how extra-linguistically uncooperative it may be (consider for instance a suspect refusing to answer a interrogator's question's in a relevant way, or, to take one of Lumsden's (2008) examples, the case where a burglar asks you the combination of your safe and you refuse to answer), will be CC cooperative, because even if the exchange is uncooperative with respect to goals that go beyond the communicative exchange, it will be minimally cooperative as long as conversational participants agree through their behaviour to maintain communication, in the sense of understanding each other. In other words, if CC is not observed, communication cannot and will not take place. CC is therefore a default feature of every conversational exchange. It cannot consequently be used as a relevant criterion in the description and explanation of instances of communication where the degree-sensitiveness of cooperation seems to play an important discriminatory role, such as persuasion, manipulation, evasive communication and the like.[10] Gu (1993) argues against the charge of triviality by assuming that if we discuss cooperation from the perspective of the talk exchange

[10]See section (§I.3.3.) for a discussion of persuasion, argumentation and manipulation.

(and not from the unilateral perspective of an individual's contribution), it does escape triviality, to the extent that when envisaged in actual interaction, cooperation describes the way conversational participants adapt their own communicative behaviour to each other's. One could in addition salvage CC by linking it to rational behaviour, which amounts to assuming that it should be construed as some kind of 'rational readiness to communicate' designed to achieve communicative goals in an efficient way. As noted earlier, it is a necessary though not sufficient condition for successful communication, thus licensing the hypothesis that it emerged as an outcome of evolution.[11]

In a recent paper, Davies (2007) provides some useful insights in this direction, by postulating that what Grice was really after was not so much an account of cooperation but one of rationality. She convincingly argues that the presence of the term 'logic' in both of Grice's original papers on implicature suggests that his main concern was to describe conversational standards, which are taken to map, in communicative behaviour, the way our mind works. As Davies notes, Grice's theoretical contribution "is not about conversations being *cooperative* – that might be an outcome of the logical structure but it is certainly not its *raison d'être*" (2007: 2321, author's italics). Grice himself acknowledged this: "one of my avowed aims is to see talking as a special case or variety of purposive, indeed *rational* behaviour" (Grice 1989: 28, my italics).

Thus, if meanings can be predicted from utterances, a systematic account of how this is possible must at the same time assume and shed light on some kind of conversational logic or architecture, itself determined by specific features of human rationality. In fact, according to Davies, Grice's mention of non-linguistic examples supports the idea that (expected) rational behaviour as witnessed in communicative interaction is not an exclusive feature of linguistic behaviour (for instance: "If I am mixing ingredients for a cake, I do not expect to be handed a good book, or even an oven cloth […]", (Grice 1989: 28)). So his discussion of implicature is meant to highlight their rational aspect, but not so much to put emphasis on cooperation. This is why Davies concludes, as already noted above, that "cooperation is the realisation of rationality applied at the level of discourse" (Davies 2007: 2324). We will see further along (see §I.1.5.) that cooperation may actually be assigned a much more fundamental role in the evolution of communication, even if this does not exclude the idea that rationality indeed seems to be fundamental to cooperation. Furthermore, to the extent that most definitions of rationality are usually expressed in terms of goals and purposes, we could see in Davies' contribution a promising convergence with the idea that CC is a feature of communicative rationality.

[11] See below (§I.1.5.) for a development of this argument on the basis of research in evolutionary psychology and anthropology.

Philosophical discussions about the properties of rationality in communication point to a few core assumptions, among which we find the idea that rationality requires from rational agents an efficient and maximising management – particularly, following Relevance Theory (Sperber & Wilson 1995) in terms of the balance between the amount of cognitive efforts it takes to process information and the effects one could expect as compensation for the effort expenditure – of the purposes, goals and means to attain these, given the cognitive and time limitations that constrain information-processing. An additional and necessary development of this would be to consider that rationality relies on different heuristics, to the extent that *ideal* rationality is cognitively implausible, which also happens to be an assumption underlying most typical objections to Grice's model. In line with Allott's (2007) work, we can extend this general view on rationality to the area of communicative rationality, which is actually an underlying feature of approaches such as Relevance Theory.[12]

My point, which gains support from Davies' remarks on Grice's philosophy, is that CC as defined above, rather than constituting a principle which actually plays a role in the derivation of meaning, is a feature of human rationality which ensures that communication *simpliciter* is possible. So whether what seems to be a shared goal (that communication successfully takes place between speaker and hearer and that both end up entertaining a similar representation) translates or not into cooperation is not so much problematic if we consider that it is inherent to human rationality in communication. Again, I have mentioned earlier that according to Davies, Grice's enterprise was to inquire about rationality more than about cooperation proper; CC could as such be taken as an attempt to ground the premises of communicative rationality:

> "Grice's interests were in the system of language as an example of human rational action, and this is to be accounted for through some variety of logic (although, perhaps, not traditional formal logic). His aim was to find the logic of conversation which could account for the gap between saying and meaning, saying and implicating, conventional and non-conventional meaning. The logic that he sought was seen as a manifestation of rational action." (Davies 2007: 2328)

To sum up, I have argued that there is room to develop in more detail the role of rationality in communication, even more so if we establish distinctions scoping over conversational behaviour in terms of the types of goals participants may be willing to pursue when they engage in communication. The first of these, which I have characterised following Chilton (1987) and Attardo (1997a), though in a much restricted sense than theirs, as the fulfilment of communication *simpliciter*, can be translated into CC. This default level of cooperation need not

[12] See Allott (2007), section 1.1.4 in particular (16-22).

be construed as a conscious undertaking, since it seems to be an inherent feature of communicative rationality, which was in part what Grice was after. However, as Davies (2007: 2329) rightly puts it, "rationality may be the driving force that Grice identified, but it does not account for how we make particular choices at the utterance level". In other words, postulating that conversational participants minimally cooperate in order to make the transmission of meaning possible does not quite inform us on how it is exactly that specific meanings are managed in communication. While the type of rationality we could expect to characterise the emergence of CC is conceivable as some kind of readiness to engage in communicative exchanges, the mechanisms involved in the actual management of specific meanings call for a more detailed account, which is what Grice was after with the postulation of conversational maxims. This is the motivation behind my own dichotomy of Attardo's locutionary cooperation and Chilton's functional cooperation into *communicative cooperation* and *informative cooperation*, which I will develop in what follows.

I.1.3. INFORMATIVE COOPERATION

While the preceding section tried to establish the relevance of CC as a rational basis for the mere *possibility* of meaning, this section on informative cooperation (henceforth IC) deals with the actual, dynamic, construction of meaning in conversation. I chose the term 'informative' because the level of cooperation I am focusing on here is the level where meaning, in terms of informative content, is managed by conversational participants. I will argue that IC is thus a different kind of cooperation, which is made possible by CC, as I take the latter to somehow carry the expectation of the former. What I call IC then maps onto those aspects of meaning involved in Chilton's (1987) functional cooperation, Attardo's (1997a & 1997b) locutionary cooperation, Gu's (1993 and 2003) pragmatic cooperation, Lumsden's (2008) linguistic cooperation, and Pavlidou's (1991) formal cooperation, which all refer to cooperation in the making, transmission and interpretation of meaning.

If we follow Austin's definition and try to reinterpret it into Gricean terms, illocution is about those aspects of meaning that correspond to the speaker's intention which the utterance is taken to fulfil. In other words, illocution is much more than just a *feature* of rationality; it refers to the way meanings are successfully communicated (that is, both produced and interpreted). So while I dissociate CC from pragmatic considerations related to the processing of full-fledged meanings, IC precisely denotes the dynamic procedures by which specific meanings are arrived at, and in this respect I posit that this degree of cooperation is concerned with actual cognitive mechanisms of meaning derivation.

It is at this stage that trends of research in pragmatics differ most, that is, in their account of meaning construction. While some approaches build on Grice's conversational maxims (Levinson 1983, 2000, Horn 1984, Bach & Harnish, 1979), others carry on with the Speech Act theoretical programme (Searle & Vanderveken 1985), while a last strand of research advocates a reductionist take on Grice (Relevance Theory for instance). So IC is the kind of cooperation where the discussion of meaning construction should take place. In a Gricean framework, this would be the level at which conversational maxims operate.

In the framework of Speech Act Theory (henceforth SAT), illocutionary acts are successful when their addressee is able to identify the illocutionary force associated to the propositional content of the utterance. Accordingly, a hearer will interpret "I'll come tomorrow" as a promise the moment he is able to identify that the speaker applied the illocutionary force of promise onto the propositional content. As such, the success of illocutions requires i) that their illocutionary force be made public, and ii) that their illocutionary force be actually recognised by an addressee. In Gricean terms, successful illocutions would be achieved the moment a hearer recognises the speaker's *intention* to communicate them. Both frameworks acknowledge the importance that the recognition of speaker intention bears in communication. And it is definitely a central feature of IC that a speaker's production bears a communicative, i.e. 'meant-to-be-communicated', status.

Austin's account of how it is that we identify illocutionary forces (relying mostly on linguistic conventions – whereby we are able to identify for instance promises by virtue of our knowledge of how promises are linguistically expressed) has been questioned,[13] but it is not my purpose to enter this discussion at this point. Rather, I'd like to emphasise the scope of IC, so as to differentiate it from CC.

I will take IC to refer to the observance by both speakers and hearers of the mechanisms by which specific meanings are successfully conveyed and processed in communicative exchanges. Following Grice, and building on the actual lexical signification of the term 'meaning', I will assume therefore that illocution is concerned with intentionality. IC, as a consequence, is about making sure that the intended speaker meaning is properly transmitted, which takes us back to Grice's conception of cooperation as a "maximally effective exchange of information" (Grice 1989: 28). If we try to produce an argument similar to (4) above, in order to address the relationship between cooperation and illocution in terms of the goals of conversational participants, this is how it could tentatively look like:

[13] See Sperber & Wilson's critique of Speech Act Theory (1995: 243-254).

(6) Informative cooperation towards a shared goal:

- A speaker who observes CC must have the goal of wanting the hearer to entertain a specific set of representations which is similar to the one she intends to communicate.
- A hearer who observes CC must have the goal of retrieving a specific set of representations, which is similar to the speaker's intended meaning (i.e., the set of representations she intended to communicate).
- Accordingly, their shared goal is fulfilled only when both interlocutors share a similar representation.
- The fulfilment of this goal is possible because comprehension follows specific rational standards of information-processing (both on the production and reception end).

So whereas the goal of CC was to ensure the mere *possibility* of meaning transmission (which can be taken as a rational stance), the goal of IC is to ensure that a *specific* meaning gets across. The former implies that conversational participants share the goal of communicating *simpliciter*, regardless of the actual point the speaker wants to make. The latter implies that conversational participants share the goal of working out specific meanings, and therefore leaves room for an account of meaning construction to be developed. IC is the stage where cognitive machinery involved in information-processing is put to work by conversational participants. Furthermore, CC entails IC in communication, to the extent that the moment an individual recognises the manifestness of a verbal stimulus, s/he will *ipso facto* try to make sense of it.

Grice's solution, which he developed in his "Logic and Conversation" ([1975] 1989), was to postulate the existence of conversational maxims which are deemed to be observed by communicators as conversation unfolds. The Speech Act Theoretical take, on the other hand, held that an account of meaning must consider the conditions under which speech acts can be deemed to be felicitous (i.e. felicity conditions). The idea postulated by Austin (1962) and developed by Searle (1969) – who himself amended Grice's account by adding a layer of conventionality in the notion of meaning – was that comprehension and, *a fortiori*, communication, is successful only when such conditions are met.

Within the Gricean framework, proper operation of conversational maxims is what makes meaning management cooperative. A speaker communicating truthful and relevant information and her addressee interpreting this information as truthful and relevant can be said to cooperate in the "maximally effective exchange of information". This is compatible with the idea that the

dynamic mechanisms involved in the transmission of meaning are cooperative, insofar as they rely on the conversational participants' respective expectations about each other's behaviour in the communicative exchange.

Regardless of the account of actual meaning derivation we opt for, dealing with IC invariably supposes that we address the actual display and use of our rational abilities, in order to explain the generation and interpretation of specific meanings. This is what Grice hoped to account for: the rational standards (i.e., some kind of *logic*) which seem to regulate conversational exchanges between humans. So whereas CC is embedded in rationality as an *a priori* necessary condition for communication, IC reveals rational processing of information; it is thus inherently dynamic as it involves cognitive mechanisms of information-processing of communicative (usually linguistic) stimuli.

An additional characteristic of IC, just as CC, is that it seems to be observed even in cases one would straightforwardly qualify as uncooperative on a non-conversational level. Take the example provided by Lumsden, where a mother replies either (7) or (8) to her son's statement "I would like to join the army" (Lumsden 2008: 1901):

(7) Mother: "I will make an appointment for you with a recruiting officer in the morning"

(8) Mother: "I have already lost a husband and two sons to that accursed army"

While (7) is by all means a cooperative answer – in all acceptations of the term – (8) does not appear to be, or at least is only partly cooperative. In (7) the mother takes on board her son's goal of joining the army, while in (8) she does not. However, this goal is extra-linguistic, as it refers to something the son wants to achieve independently of the conversation he is having with his mother, and that he can tentatively achieve without relying on his mother. Now, in both (7) and (8) the mother inferred correctly what her son meant by formulating his utterance. And, even if in (8) it could be said that the mother is not cooperating to fulfil her son's goal, as she refuses to support his joining the army, she can definitely be said to be IC-cooperative, since she has interpreted her son's utterance, and correctly inferred his intention to communicate that he wants to join the army. IC has to do with the actual instantiation of a specific meaning, from its production to its interpretation. And it is not necessary that interlocutors agree with, or adopt each other's extra-linguistic goals to make sure that they understand each other. That is why IC is limited to matters of meaning construction.

Furthermore, even in cases where hearers have strong assumptions about the speaker's untrustworthiness, they will still be able to recover speaker meaning. The fact that I may not trust such and such politician when they make public promises on TV does not prevent me from

understanding the meaning of their utterance, and that they are performing a promise, even if I may doubt that they will abide. The same goes for the interpretation of metaphors, which blatantly violate the maxim of quality. It is not because I don't trust a politician that I will systematically doubt any of his ostensive maxim violations. I will still be able to retrieve the intended meaning, at a representational level, simply because entertaining a representation and entertaining a belief as true are two distinct things. In other words, understanding is not the same as believing, though it seems quite straightforward to assume that the belief that P is true requires the individual's prior understanding of P. In any case, this is further evidence to point out that IC operates the moment someone deems a stimulus worth of interest. To this, we could add that IC operates the moment CC operates.

I have assumed CC to be a feature of rational communication. IC points to the fact that conversational exchanges are somehow rationally structured, and that they rely on stable cognitive mechanisms which turn thoughts into propositions for purposes of communication. Once we acknowledge this idea, several accounts of how this rational structure works and is organised may be confronted and compared in order to find the most satisfying and cognitively plausible solution. The Gricean model has been attacked by many people, some of which making a real point. It will be the object of another section to expose the details of these critiques of the CP and to provide an alternative.[14] However, this is not my point for the time being; I so far merely intended to spell out the idea that the unfolding of meaningful exchanges can be described as an informatively cooperative undertaking. In this sense, my bipartition of Chilton's cooperation for communication *simpliciter* and Attardo's locutionary cooperation into CC and IC is a further clarification of what Grice's CP can be taken to mean on the level of the exchange of meaning; they are consequently not meant as a critique of Grice but rather as a means of making explicit some implications of the cooperative conception of communication.

Let us now turn to the characterisation of what Attardo (1997a, 1997b) calls *perlocutionary* cooperation, which I have briefly evoked in what precedes.

I.1.4. PERLOCUTIONARY COOPERATION

A widespread view of cooperation, which also builds on the everyday meaning of the term – and which Davies takes to be responsible for the "cooperative drift" (Davies 2007: 2314), or the amalgamation of two different acceptations of the term *cooperation* – usually considers it to be geared towards the fulfilment of extra-linguistic goals. In fact, in the section he devotes to the

[14] See section on Relevance Theory (§I.1.7.)

discussion of examples of implicature (Grice 1989: 32), the first case Grice examines – reproduced below in (9) and (10) – is an illustration of such an extra-linguistic goal:

(9) A: I am out of petrol

(10) B: There is a garage round the corner

Here, the implicature A may derive from B's utterance, namely that the garage is open and that it must sell petrol is reached by assuming that B is observing the maxim of relation. Furthermore, in terms of cooperation, A is entitled to infer that B's contribution is *relevant to the goal* A is pursuing in making her utterance, namely her *extra-linguistic goal* (to fill her tank with petrol). It is on this level of cooperation, which Chilton terms "communicative co-operation, corresponding to wider ethical norms" (Chilton 1987: 225), that Grice has been most attacked, since there are grounds to doubt that all conversations are geared towards extra-linguistic cooperation. The argument is quite simple: we witness in everyday conversations, particularly in situations where interests between speaker and hearer conflict, that conversational participants do not work jointly towards a shared goal. This is because extra-linguistic goal adoption does not obtain in those particular situations. However, this does not have a direct consequence on CC, nor on IC, because successful communication, regardless of its social outcome and consequences, supposes that conversational participants share the goal of understanding each other, i.e., in Attardo's terms, the goal of "maintaining communication" (Attardo 1999: 461).

Most authors (e.g. Chilton 1987, Gu 1993, 2003, Attardo 1997a, 1997b, 1999, Pavlidou 1991, Lumsden 2008) who have worked on the CP after Grice have come to the conclusion that his formulation left room for a double interpretation of cooperation, the first of which we have dichotomised in the previous two sections. The second acceptation of cooperation, which I will call following Attardo (1997a, 1997b) *perlocutionary cooperation* (henceforth PC), corresponds to Chilton's (1987) *ethical cooperation*, Pavlidou's (1991) *substantial cooperation*, Gu's (1993) *rhetorical cooperation*, Lumsden's (2008) *pursuit of extra-linguistic goals*, and basically builds on the idea that communication is a means by which human beings can satisfy goals which go beyond the linguistic exchange. As shown in the example above, PC is about i) acknowledging and adopting the speaker's extra-linguistic goal and ii) making one's contribution helpful (relevant, appropriate) with regard to that extra-linguistic goal.

PC bears a moral aspect which both Chilton and Attardo recognise. Chilton refers to it as "cooperation corresponding to wider ethical norms" (Chilton 1987: 225), and Attardo, in his formulation of the Perlocutionary Cooperative Principle (PCP), states that conversational participants should act as "good Samaritan[s]" (Attardo 1997a: 766). The incorporation of a moral component in a pragmatic account may raise some difficulties, to the extent that it might

prove problematic to put such a requirement at interplay with cognitive mechanisms of information-processing. The relativity involved in ethical considerations may be too much of an obstacle for a naturalistic account such as the one I am after; [15] if PC depends on moral norms, then we must be prepared to accept that these may be culturally-bounded, and therefore that they may vary from one community of speakers to another. In the theoretical research I am currently involved in, I chose not to explore this track, although I am aware that such a discussion might come as a necessity at some point, notably when it comes to assessing whether cultural norms influence the amount of cooperation conversational participants are led to observe. But that is not the purpose of this research, and the project of discussing these points falls outside the scope of linguistics.

In order to elaborate a consistent model in which CC, IC and PC could coherently be accounted for, it would be interesting to integrate the outcome of evolutionary psychology and anthropology research, whereby cooperation is considered to be an advantageous behavioural option for the survival of a species, as their conception of cooperation precisely addresses goals individuals may be wanting to attain beyond communicative exchanges. This would confer a sociological tone to the account, but would as well leave the door open to the idea that the kind of ethics or morality involved in human interaction (in particular in communicative exchanges) is to some extent a natural feature of cooperation, even if the way it would express may turn out to be contingently conventional. It should also be noted at this point that such a discussion need not presuppose that this must be expressed in terms of a rigid normative system: the fact that we are often confronted to situations where people deceive others would easily disqualify such a claim. For the time being, I will assume that PC can be envisaged as some kind of default assumption about what type of behaviour people generally expect from others in conversation and that deception precisely exploits our natural predisposition to perlocutionarily cooperate. These issues will be elaborated in Chapter III later on.

Coming back to the discussion about the features of PC, it seems that although it is aimed at regulating behaviour exceeding the scope of a communicative exchange, perlocutionary cooperation, as Lumsden (2008) puts it, may have a role in the construction of meaning (notably implicatures) and the generation of an appropriate response to a communicative stimulus. To take his example (Lumsden 2008: 1904), the same information (for instance information about the quality of a car) supplied by a salesperson or by a close friend may be interpreted very differently; in the first case one would probably associate a lower degree of plausibility to the

[15]But see Attardo (1999) for an attempt at a formalisation of these constraints in terms of goals and subgoals of conversation.

truth of the proposition, since it is obviously manifest that the salesperson is pursuing the goal of selling a given product and that s/he may be untruthful in view of fulfilling that goal, while in the second case, the degree of plausibility of the proposition may be interpreted as higher, in that we usually trust our friends and therefore assume that they will be cooperating towards whichever extra-linguistic goal we may want to pursue.

Now, both cases display CC and IC, since communication is successful and meaning has been conveyed. However, the actual output representation will vary depending on the amount of PC we may ascribe to the speaker. Lumsden defends the idea that PC may thus determine the interpretation of the stimulus in cases where there is a manifest – and mutually recognised – conflict of interest between speaker and hearer. Although Lumsden's analysis makes the interesting point that the outcome of communication may vary depending on PC, it fails to make a distinction which is nonetheless crucial: while in one case we ascribe higher plausibility (when the information is uttered by a close friend) and in the other lower plausibility (when the information is uttered by a salesperson), in both cases the 'informative' meaning is the same; whether a salesperson or a close friend says it, the utterance means the same. In terms of speaker meaning, then, the content of the utterance is similar, regardless of the amount of PC the speaker puts in the conversation. In other words, what Lumsden's example shows is that *believing* may depend on PC, not that *understanding* does. The question of knowing why information will be judged reliable is in principle distinct from the question of knowing how people understand information. In the example just mentioned, whether the information is uttered by a salesperson or a close friend does not affect its understanding, but rather the attitudes that the hearer may adopt as a result.

In any case, his account of cooperation may provide insights on cases where the speaker is typically assumed to be uncooperative; the idea being that extra-linguistic uncooperativeness correlates to the absence of shared extra-linguistic goals. Ultimately, his take confirms the expectation of rational cooperation in communicative exchanges, but also implies that conversational participants may process information under the assumption of a rationally limited PC when conflicts of interest arise between them. It should also be noted in this respect that a breach in PC does not necessarily entail a breach CC or IC since the pursuit of extra-linguistic goals, in particular situations, does not affect otherwise rational expectations of communication (in particular the assumption that language users who communicate need to make sure that information is properly exchanged).

This view, which considers the actual interaction as a unit of analysis (somewhat in Gu's (1993) vein), shows that because such breaches in cooperation may occur, PC may be seen as an essentially dynamic notion, and that the scope of cooperation (i.e., the goals which cooperative

conversational participants aim to attain) may change potentially every time a speaker makes a contribution. Take examples (7) and (8) again. In (7), the mother's contribution is cooperative with respect to her son's goal, which is to eventually join the army. In (8), her answer is uncooperative with respect to such a goal, but at the same time calls for the son's cooperativeness towards the opposite goal, namely that he doesn't join the army, or at least makes room for such a discussion to take place. Whether the son cooperates or not, in the end, is altogether a very different question; nonetheless, in uttering (8), the mother communicates that she will not cooperate with her son's goal and at the same time that instead, she instantiates an alternative goal. What this suggests is that it only makes sense to speak of PC *with respect to a particular extra-linguistic goal*. Furthermore, it indicates that cooperation dynamically evolves as speakers take turns in reacting to each other's contribution. So while we usually speak of cooperation as describing the overall conversational participant's behaviour in absolute terms, a more global level of PC, that of discourse, would seemingly be more difficult to grasp. Insofar as successful conversational cooperation towards a goal requires both mutual recognition and acceptance, and given that the goals speakers pursue and make manifest evolve as communication unfolds, it becomes very difficult to assess such a global level of cooperation.[16]

For the time being, let us proceed with the discussion by considering in what follows an evolutionary perspective on cooperation, so as to assess an integration of these three levels of cooperation in a wider framework. This will allow me to consider more precisely the differences between the three identified levels of cooperation, and perhaps even lead me to question whether the first two are indeed instances of cooperation.

I.1.5. EVOLUTIONARY BASES OF COOPERATION.

According to Clément (2006), who draws on Dunbar (1998) and paleontological research, at some point hominids were forced to leave forests, which were becoming far too hostile environments, and migrated to the savannah. A radical change of environment usually prompts for adaptive changes in the species; perhaps one of the most spectacular adaptations was the adoption of the biped stance, which supposed a great survival advantage, in that it allowed spotting predators from far and accordingly allowed enough time to take appropriate action to

[16] Some researchers such as Sperber and Wilson (1995) implicitly contend that PC is not at all relevant to explain how communication works and they reduce it to a contextual assumption as any other contextual assumption taking part in meaning construction. Yet, this would accommodate nicely with the idea, mentioned earlier, that PC may evolve throughout conversations. A construal of PC as a contextual assumption the hearer mobilises as he processes the speaker's utterance may actually avoid these difficulties.

avoid danger. However, bipeds were still fragile and vulnerable creatures – relatively to many other species which could outperform them in strength, stamina, and, consequently, chances of survival – that needed to feed and therefore both hunt and gather. It is nowadays widely accepted that primitive social organisation emerged as an effective way of maximising chances of survival, in so far as it is easily conceivable that cooperative action may outperform individual action in those scenarios.

Life in communities of individuals, what is more, in communities under pressure to survive, can thus considerably benefit from some kind of organisation, in particular with respect to activities which one individual would have little chance to succeed in alone. In this respect it could be argued that hominids would gain great advantage from designing some kind of collaborative system in order to hunt powerful and stronger preys, so as to ensure food supply. And therefore cooperative action emerges as an interesting and possibly advantageous option, firstly by virtue of the positive results it yields (in comparison to unilateral action – one individual would probably go through significant trouble in killing, say, a powerful bovid, while a group of individuals acting together may succeed more easily) and secondly because its outcome usually turns out to yield a great benefit for all cooperators and *a fortiori* for the community they live in.

I.1.5.1. The prisoner's dilemma and reciprocal altruism

Over the past century, many scholars have tried to formalise the problem of cooperation, some of them with striking results in the framework of game and decision theory. In this line of research, issues on cooperation have revolved around the so-called prisoner's dilemma (and its subsequent development, the iterated prisoner's dilemma, cf. Axelrod [1984] 2004), which is formulated as follows: imagine a game with two participants who find themselves in a situation where they must interact, and who may do so either by cooperating or defecting. When both cooperate, they get a shared benefit. If one defects while the other cooperates, the defector gets a greater benefit than if he had cooperated, and the cooperator gets nothing. If both defect, they still get a benefit, but lower than the one they would have gained by cooperating. This is summarised in figure 1 below (inspired from Axelrod 2004):

A \ **B**	**Cooperate**	**Defect**
Cooperate	3;3	0;5
Defect	5;0	1;1

Figure 1. The prisoner's dilemma

Now, let's take the perspective of one of the participants, A, and consider his options, based on the behaviour his co-participant, B, may engage in. If B cooperates, then the most beneficial option for A is to defect, because A will get 5 units reward instead of 3, had he cooperated. If B defects, the most beneficial option for A will also be to defect, because he will get 1 unit reward, instead of 0 if he cooperates. So it seems that whatever B's behaviour turns out to be, it will always be better for A to defect than to cooperate.[17] Yet, looking at the overall outcome of the conjunction of behaviours, the table also indicates that participants would be better off if they both cooperated instead of both defecting.

The outcome of the prisoner's dilemma (i.e., the preference for defection) is somewhat puzzling if we consider quite straightforwardly that humans are generally more likely to cooperate with each other. The story is nevertheless not as simple as one single instance of the prisoner's dilemma: this line of reasoning holds only if the game is played just once. Indeed, in case the game is known to be a one-off situation, then defection is most likely to be the preferred option, since there would be no tangible negative consequences for defection, given that only one instance of that particular game is played. However, the one-off version of the prisoner's dilemma is perhaps not the best illustration of what could possibly happen 'in real life'.

Humans, even more so if they are members of the same social group, are likely to interact with the same people more than once. This is why we must further consider the *iterated* prisoner's dilemma, which denotes a multiple-round game, where participants meet over and over again. In such a context, any action will be likely to bear consequences on further interactions between A and B. This is so to the extent that we need to take into account that both are able i) to recognise each other in future exchanges, and ii) to remember how each of them behaved in previous rounds. So if A defected while B cooperated, there are chances that in the next encounter between A and B, B will be suspicious of A and as a consequence might be unwilling to cooperate, by fear of being tricked again. An additional feature of the iterated prisoner's dilemma, which does not apply to the simple prisoner's dilemma, is that it increases the chances of emergence of some kind of punishment for not cooperating. This is particularly true if we consider that the game can apply to any real-life situations where people are led to meet over and over again. It is thus a first indication that defection in the long run may not be advantageous for an individual: he faces potential punishment, which could translate into rejection from his co-participant, and tentatively from the social group to which both belong.

[17] Notice also that these conclusions can be drawn only relatively to B's behaviour. This suggests that any account of cooperation must take into consideration that the cooperative/defective behaviour makes sense only relatively to someone else's behaviour and not in absolute terms.

A further layer of complexity must be mentioned here: if the game has a mutually known fixed number of rounds, then participants will know that, as far as that particular game is concerned, there will be no punishment after the last round, which could encourage defection at the last stage. In turn, this thought may encourage defection at the stage before last, since co-participants will be able to foresee that defection is likely to obtain in the last round, and so on and so forth until the very first stage. Now, it must be acknowledged that in real life, we seldom, if never, know how many times a 'cooperative game' will be played, or, more generally, how many times we will be led to interact with the same person. Therefore, ignorance of the game's duration proves to be an additional relevant parameter and might indeed determine the outcome of participant behaviour in the game. In other words, given the uncertainty of the number of times the game will be played between A and B, participants would be better off cooperating, since i) it secures a better outcome for both than mutual defection, and ii) it increases the chances that the other participant will cooperate in the next round, instead of defecting (thus avoiding punishment and retaliation).

Now, even if the prisoner's dilemma model is an artificial one, it could be argued that such situations can occur in the biological, natural, world. Members of a same (relatively small) group are likely to interact time and again and in situations in which their behaviour may have beneficial or detrimental consequences for one another or for the group. In principle, thus, when two members of the same community meet, we should expect to see behaviours such as cooperation, defection, or exploitation (in terms of deception). Furthermore, assuming that the members of the community are equipped with cognitive abilities that may allow them to recognise conspecifics and to remember past behaviours, situations where cooperation is at stake are seemingly likely to be also encountered in animal interaction.

Mathematician and political scientist Robert Axelrod (cf. Axelrod [1984] 2004) was one of the first to test the prisoner's dilemma on a large-scale computer simulation in order to experiment the effectiveness of different strategies in cooperative/defective scenarios. What he did, in the 1980s, was to invite researchers in game theory to submit computer programmes designed to compete in a prisoner's dilemma tournament. Among all the strategies that were implemented by software designers, the most effective – and this happened repeatedly – was Anatol Rapoport's. His code was extremely simple: it began with default cooperation in the first round; then, it would behave according to the co-participant's behaviour, by adopting the same behaviour the latter engaged in, that is, punishing him for defection (i.e. defecting), and rewarding him for cooperation (i.e., cooperating). This strategy was called the Tit-For-Tat strategy, and the programme implementing it won the tournament. In the end, it showed that cooperation was the long run preferred strategy, since it enforced systematic punishment for

defective behaviour and systematic rewarding for cooperative behaviour. Thus, it would elicit cooperation by acting on the co-participant's behaviour. This is actually a powerful feature of the Tit-For-Tat strategy: it ensures that more collaborators, or reciprocators, will join the gaming sequence over time, while at the same time eliminating defectors.

A couple of remarks on this particular strategy should be made here. First of all, it shows that cooperation does not require trust nor friendship, but rather that both participants conform to the strategy over time. Axelrod (2004) mentions the example of the "Live and let live" strategy, where soldiers from opposing camps in World War I would refrain from attacking their enemies' supply of food, and sometimes from shooting to kill, so that their opponents would do the same. And it actually initiated some kind of cooperative dynamics for a period of time, very much similar to the iterated prisoner's dilemma: enemies began to cooperate tacitly, knowing that failure to cooperate would be detrimental to both camps, since an attack would inevitably yield further reciprocal retaliation. The reason why, according to Axelrod, this system worked is because both camps did not know for how long they would be fighting each other on the front. In addition, it can hardly be assumed in this case that enemy soldiers trusted each other, let alone that they were engaged in a friendly relationship. The crucial condition of the strategy, therefore, seems to be the uncertainty about game's duration, rather than trust among participants.

Another issue one might bring forth is the status of rationality in the prisoner's dilemma. Of course, acting to ensure one's own benefit could easily be considered as rational action, in that it involves consideration of a goal and the means to attain it efficiently. However, the Tit-For-Tat strategy does not require rationality to function, and this is a point Axelrod explicitly makes: "there is no need to assume that the players are rational. They need not be trying to maximize their rewards. Their strategies may simply reflect standard operating procedures, rules of thumb, instincts, habits, or imitation" (Axelrod 2004: 18). All the game requires is that both participants follow a strategy, so that the interaction develops quasi-naturally, somehow with an inertia-like dynamics, following a simple stimulus-response pattern (leaving aside the initial cooperative move). Nevertheless, one might still wonder whether the adoption of a strategy is the fruit of rational calculation or not. If we trace back to the evolution of cooperation among social groups in time, there are grounds to hypothesise that, *at the very beginning at least*, the processes that led to the adoption of such a behaviour were somehow rational, in that *someone had to understand that it was to their advantage to adopt the Tit-For-Tat strategy*, or, in other words, that cooperation in the Tit-for-tat sense was actually instrumental to achieve certain goals. This last conception supposes that we conceive of rationality as practical reason, in the sense that something is rational if it is instrumental to the fulfilment of a specific goal. If we take

practical reason as a frame in which cooperation could have emerged and developed, then it does make sense to assume that it constitutes rational behaviour.

One of the main theories that emerged to explain why cooperation proved to be an advantageous option through evolution hinges upon the prisoner's dilemma; it is called reciprocal altruism (Trivers 1971, see also Axelrod [1984] 2004). Under this view, and as noted above, members of a consistent social group will be likely to cooperate in the long run, because i) they can remember each other and their respective behaviour, ii), they may accordingly avoid punishment and iii) they secure that the game will be played again (allowing for further exchanges). Engaging in a reciprocally altruistic behaviour seems thus a viable option to the achievement of those goals.

So we are left with the idea that cooperation may have emerged in order to foster successful and efficient social group life. As noted at the beginning of this section, it proved viable an option for early hominids; according to game theoretical research, the Tit-For-Tat strategy provides a suitable dynamics of cooperation, since it ensures consistency in the social group (through reciprocal altruism). What this seems to indicate, all in all, is that cooperation may very well have resulted from social requirements for community organisation.

I.1.5.2. Communication, language and cooperation in evolution

I am now able to link cooperation to the emergence of communication, so as to combine (and hopefully benefit from) an evolutionary perspective and Grice's work on the CP. Here is a first take on how the argument could go: if we consider that cooperation, in the long run, is an advantageous option for the survival of the species, then it follows, all other things being equal, that the more effective cooperation is, the better off the species will be. And one way, though not mandatory, of making cooperation effective is to adopt a system which will be able to ensure efficient transmission of information (for instance the ability to inform about preys, predators and tentative dangers, as well as the ability to give orders may secure fruitful course of action, in the case of group hunting). This is arguably a sensible hypothesis which can explain the emergence of communication as an appropriate system of interaction designed to maximise cooperation. It does make sense to assume in this vein that communication emerged as a successful and convenient way of maximising cooperation.[18]

[18] This is what Hart (2010: 37) argues for when he considers the issue of "communication for cooperation".

However, all the debates on cooperation around the prisoner's dilemma revolve around PC, since the goals towards which participants of these games tended were presumably extra-linguistic goals: in the case of WWI soldiers, securing food supply or refraining to shoot to kill, in the case of early hominids, hunting, and in the abstract case of the iterated prisoner's dilemma, getting a symbolic reward.

This, I contend, is because *communicating* in order to facilitate cooperation is not exactly the same thing as *using a language* in order to facilitate cooperation. The difference is in my opinion very relevant, for being able to communicate is quite different from being able to use a language (see Sperber 2000a and a discussion in the next paragraphs), and I argue that, in evolutionary terms, this distinction could allow us to ground the aforementioned distinction between CC/IC on the one hand and PC on the other. In short, such a distinction points to considering language and communication as two separate but related *loci* of cooperation.

Cooperation does not necessarily require communication, as Axelrod's work convincingly shows. But this says nothing about whether communication may maximise cooperation or not. As a matter of fact, ordinary situations allowing communicators to share information (be it verbally or otherwise) about their environment and the courses of action they could be willing to engage in can foster cooperation, by virtue of the emerging possibility of organised collaborative behaviour, be it planned in advance or decided on the spot. Nonetheless, this still does not require language, or to put it more precisely, recourse to a structured and stable system of symbols and referring units.

In his work on communication, language, and their respective cognitive underpinnings, Sperber (2000)a argues that the existence of a conventional code is not a necessary condition for communication. Intuitively at least, this finds confirmation in the idea that one can communicate by spontaneous and unconventional or novel gestures. There is no need for a shared code for that. But communication does require the communicators' ability to recognise what each other may be thinking, intending, planning, and so forth. In other words, communication requires a *metarepresentational* ability, that is, the ability to form representations about other representations, either one's own or others'. Sperber assumes metarepresentation to have emerged before language, and therefore before verbal communication. He argues quite convincingly that these are two independent abilities that have nonetheless co-evolved;[19] evidence of metarepresentational features of reasoning which are independent of language

[19] Sperber argues that if language had emerged before the metarepresentational ability, it would have had to be reduced to a coding-decoding scheme. If such had been the case, and since our contemporary communicative ability is not of the coding-decoding type, it means that a sudden change would have had to happen at some point in evolution, which is unlikely. See Sperber (2000a) for the discussion.

flourished in research on Machiavellian intelligence (Byrne and Whiten 1988; Whiten and Byrne 1997) and could be mentioned to further the claim.[20]

Hurford (2007) considers the problem in a similar vein as Sperber's, as he hypothesises that a necessary component of communication is the potential for illocution, that is, its potential to *do something* in the world, somehow independently of the existence of a shared code. In Austin's original (1962) account, illocutionary force is about the identification of the full-fledged meaning of an utterance, namely an indication leading to the identification of the intention the speaker meant to fulfil by directing a particular utterance to an audience. Illocution is intimately linked with intention, since to understand the illocutionary force of a speech act, one needs to be able to form an assumption about the speaker's intention. According to Hurford (2007: 170-177), the ability to signal something to a conspecific underlies all instances of communication. In other words, and from an evolutionary perspective, communicating, prior to involving reference and propositional content, boils down to being able to make one's intention manifest and interpretable, thus calling for a hearer's speculation about the speaker's mental states.

Communication does not necessarily require language; moreover, what seems to be central to communication is not the code (a code-model, as shown by Sperber & Wilson 1995, cannot explain the possibility of implicit meaning, which is contextually inferred), but the ability to make manifest and recognise intentions, which in turn makes for inferential communication, as defended by Sperber (2000a). Now the interesting point is that, pretty much the same way communication may make cooperation more effective, language may make communication more effective, in that the existence of a system that makes reference possible while at the same time taking advantage of inferential processes of meaning management is a serious advantage when comes the need to engage in collaborative and cooperative behaviour. In Sperber's words, "metarepresentational sophistication allows a form of inferential communication independent of the possession of a common code. This type of inferential communication, however, *can* take advantage of a code." (Sperber 2000a: 6, my italics)

Leaving the discussion on metarepresentation and language aside, I assume that the emergence of linguistic competence is related to the need for communication, which is itself very much related to the need for cooperation. For the purposes of the present chapter, I do not intend to

[20] It has been for instance showed that primates are able to adopt deceptive behaviour, which indicates that they are able to foresee potential intentions and behaviours on behalf of their conspecifics. The idea being that, still without the type of linguistic communication humans nowadays master, primates are able to attribute intentions to others, so as to deceive them, or conversely protect themselves from deception. I will leave out of the discussion whether or not this constitutes an indication that primates possess a proper theory of mind.

enter the discussion about strict causal relationships between these, and will therefore limit myself to noting that there seems to be a sensible correlation between their respective evolutions. Still, there are reasons to consider that CC and IC (which are directly related to the possibility of language) maximised the efficiency of communication, which itself was able to maximise *extra-linguistic* cooperation, as an advantage involved in survival. On this precise issue, Origgi & Sperber state that "metarepresentation and communication are […] necessary precursors for the ontogenetic development of language, which is adapted to meet the communicative needs of humans" (Origgi & Sperber 2000: 141). Metarepresentation and communication are necessary conditions of language; in that sense, their emergence couldn't but facilitate the development of the latter.

Pushing things a bit further, we would end up with the hypothesis that it is the need for PC – which is relevant to communication (but not directly language), insofar as it involves both purposes and goals which go beyond the linguistic (or non-linguistic) exchange of information – that facilitated the development of CC and IC. It does not matter at this stage whether the relationship between these types of cooperation is causal or not; rather, the need for the former set the conditions for the integration of the latter, even if it is perfectly conceivable that these may have developed independently.

I am now tempted to ground a distinction between CC and IC on the one hand, and PC on the other, on the basis of these considerations about language evolution and communication and the role they may have had on cooperative behaviour. It could indeed reasonably be argued that the stabilisation of communication as a way of satisfying goals that go beyond communicative exchanges has facilitated the emergence of CC and IC as well, with all the changes it involves in both the physical apparatus used to this end (mouth, throat, etc.) and the cognitive abilities the use of language requires (reference, symbolic use of expressions, metarepresentational ability, etc.). This of course implies that the need for PC was prior to the need for CC and IC, but by no means does it contradict the assumption made earlier, according to which communication emerged as a way of maximising PC. It could also be the case that CC, over time, became progressively integrated to some sort of instinctive rationality for communication. Under this view, CC would characterise the human natural *tendency and readiness* to i) intend to communicate meaningful contents, and ii) interpret verbal material, which virtually all scholars in the study of communication agree today is completely automatic.

It is here worthwhile recalling Hurford's discussion (Hurford 2007: 320-322), which draws on research on young pre-linguistic infants to argue that babies have a natural tendency to imitate stimuli coming from their parents: this could be an indication that there is something instinctive

about the readiness to communicate, leading to the hypothesis that this constitutes an evolutionary outcome. Yet, Hurford continues,

> "when a human infant starts to imitate first words, those words are not used in any practical context like carrying a log with its mother. […] Do we want to say that the infant **understands** what its mother is trying to do, and then cooperates in achieving the shared goal? I don't think so." (Hurford 2007: 320, author's bold)

So even if infants do not master language – and are consequently not able to manage propositional content in the semantic sense simply because they have not yet acquired language at that stage – they still display readiness to make something out of the stimuli they are exposed to, and this may very well indicate that they are geared towards CC 'by default' (that is, to engage in communicative behaviour that will, over time and after language acquisition, prove meaningful).

The case of IC is more problematic. While it could be tempting to assume in the preceding vein that IC is also embedded into some kind of instinctive drive, recent research on pragmatic competence (e.g. Noveck 2001) convincingly shows that it is not until a certain age (somewhere around the age of 6-8) that children are able to master full-fledged meaning, in particular implicatures. In Austinian terms, one could say that children acquire full-blown illocutionary mastery over time. This is why the distinction between CC and IC makes sense in my view: the former relates to the possibility of the exchange of meaning (the communicators' default readiness to engage in meaningful exchanges, which seems to be an innate feature of the human mind) and can be related to a natural instinct, whereas the latter is about encoding and retrieving specific meanings, which is the stage where cognitive mechanisms involving recognition of intention combine with the management of propositional contents to yield full-fledged meaning, both in production and reception. This should come as no surprise, given that language is acquired over time in the development of the infant. Consequently, IC can be said to rely both on the metarepresentational ability and on language acquisition which allows for management of specific referring propositional contents.

Where does this discussion on evolutionary aspects of cooperation, communication and language leave us? Let me summarise the main points of the section:

- PC took advantage of CC and IC to operate more efficiently, which was an adaptive requirement.
- CC can be associated with an instinctive readiness to communicate; it is a feature of communicative rationality, and a necessary condition for communication.

- IC supposes both the acquisition of language and metarepresentational ability; it therefore emerges over time, in the course of infants' development.
- the types of cooperation involved in CC, IC and PC are different, even if they may all play a role in the derivation of specific meanings.

To take on that last point, what we have seen in this section points to a clear difference in nature between the three kinds of cooperation we are dealing with. It is the purpose of the next section to elaborate on these differences.

I.1.6. SUMMING UP: THE COOPERATIVE STATUS OF LOCUTION, ILLOCUTION AND PERLOCUTION

If we think in terms of the nature of cooperation, which supposes at least two individuals sharing a common goal and undertaking joint action with the perspective of gaining some benefit from it, there is room to discuss whether CC, IC and PC involve the same mechanisms.

Let us take CC first. I have ventured that a readiness to communicate together with expectations of meaningfulness could be taken as a rudimentary kind of cooperation if we consider that people work jointly to fulfil the goal of understanding each other; under this view, such a disposition can be characterised as rationally cooperative. Conversely, a child deliberately blocking her/his ears and singing loudly at the same time so as to avoid being exposed to verbal communicative stimuli is precisely not cooperating, in the CC sense, because s/he is denying the mere possibility of communication, let alone the understanding any specific meaning.[21] This would amount to blocking one's natural readiness to communicate. As such, it is indeed uncooperative, and it amounts to denying communication *simpliciter*. But CC does not require an active stance from the individual, simply because, as pointed out before, it corresponds to a natural instinct to communicate in addition to being completely automatic (it proves difficult not to process a linguistic stimulus which is addressed to oneself, unless one deliberately makes an effort not to perceive it properly, like the child blocking his ears in the example above). Nonetheless, it can still be argued that CC requires some kind of goal sharing, even if the latter corresponds to something as automatic as engaging in communication. The interesting point here is that CC does not require the addressee to make mutually manifest that he is agreeing to cooperate through any type of confirmation. In ordinary contexts, I usually address someone without needing them to tell me "you may address me, I am now ready to communicate with

[21] Even if by doing so, he is still communicating something, namely that he does not want to engage in a conversational exchange with her/his interlocutor. Nonetheless, with respect to the latter's goal of attempting to communicate, we can still say that the child is not cooperating. Recall here that it makes sense to speak of cooperation only in relation to a specific goal.

you". This resonates with the claim made earlier, according to which this minimal degree of cooperation is somehow assumed by default in any communicative event, as part of what I called a rationality of communication.

IC, on the other hand, is more than a mere state or disposition, in that it requires cognitive processing of linguistic stimuli, and in this sense it is a dynamic process. Since it is the stage where actual, i.e. verbally instantiated, meanings are produced and retrieved, it does suppose that conversational participants make processing efforts. This entails that there is an additional layer of involvement in the success of the exchange from conversational participants. IC builds on CC, and this makes the latter a necessary condition for the former. It would indeed be counter-intuitive to assume that hearers could automatically process information at all were they deprived of a natural tendency and cognitive capabilities to do so.

One may ask at this point how this would apply to cases of overhearing, where someone processes information which is not directly intended for them. In such cases, it could be argued that there is no cooperation, on the grounds that the speaker does not have the goal of having that particular overhearing addressee (whose presence the speaker might not even be aware of) understand her utterance. Indeed, there is no cooperation in this case, but this is a mere consequence of the fact that conversational participants who have access to the verbal stimuli here are not taking part in the same conversational exchange, and in that sense it makes whatever meaning is conveyed independent from a cooperative undertaking *between those specific participants*. Still, the speaker in that case is CC and IC cooperating with *some other* hearer (the one she is actually addressing), so it could still be said that she is cooperating, albeit with someone else. The overhearing individual is merely picking up a cooperative attempt which is not directed at him. In other words, we may speak of strict IC only in cases where there is a mutually manifest communicative interaction between speaker and hearer, but this does not prevent overhearing and interpretation from taking place, and we can still say that the speaker is engaging in cooperative behaviour as she formulates her utterance. Additionally, this could indicate that a CP may not explain all cases of successful information transmission, if we take into account the fact that conversational participants may not find themselves in an actual communicative configuration. We will see further along that this idea may converge with Post-Gricean accounts of communication which refuse to postulate the necessity of a CP.

As for PC, in addition to engaging both participants in effortful cognitive processing, it requires the hearer to make manifest that he adopted the speaker's extra-linguistic goal, for instance by producing an appropriate verbal response or engaging in a relevant course of action. If I ask my partner to pass me the salt, thereby prompting for her cooperation with regard to the goal I have of adding salt to my food, the latter will only be fulfilled if she actually does pass me the salt. Of

course, if I say "Can you pass the salt", and she replies "Yes", while continuing to eat as if nothing had happened – as people sometimes do to generate humorous effects –, she is cooperating, at least on the CC and IC levels,[22] but certainly not on the PC level. Communication will certainly have taken place, there is absolutely no question about that; both of us have made manifest certain meanings and interpreted them. But PC requires from the addressee a relevant response that will be instrumental to the fulfilment of the speaker's extra-linguistic goal. This is the main difference between CC and IC on the one hand, and PC on the other hand. The first two require that the shared goal of getting the message across is pursued by both speaker and hearer; however, they do not strictly require the hearer to provide a response which will confirm that they are indeed cooperating. PC, on the other hand, is not necessary for communication to take place, simply because it presupposes a communication-independent (extra-linguistic) goal in nature; nevertheless, when such a goal is instantiated in a conversation, its fulfilment requires the hearer to make manifest in some way or another that he will be cooperating (i.e. that he acknowledges that he took on the speaker's extra-linguistic goal). Furthermore, the first two may be successful without this confirmation, while the third would not. This is probably one reason why Relevance Theorists do not consider the former to be cooperative in essence, but I will return to this shortly.

In terms of the relationship between the three kinds of cooperation operating within communicative exchanges, successful PC requires both CC and IC, for one wouldn't be able to cooperate towards an extra-linguistic goal instantiated in a communicative exchange if one is not in the first place i) willing to engage in the communicative exchange and ii) able to understand the speaker who formulates that she is requesting help to attain that goal. But the relation of necessity does not hold in the other way: CC and IC do not require PC to obtain; as such, *they are definitional of verbal communication*, while PC is not.

It is nevertheless worth noting that within a neo-Gricean framework, the three kinds of cooperation I exposed in this section bear a strong relationship with the management of meaning in conversation from the perspective of both production and reception; while it is almost trivial to assume that CC and IC are essentially involved in the recovery of meaning, Lumsden (2008) showed that the degree of PC a hearer assumes from a speaker may just as well

[22] One could question whether this is indeed CC and IC-cooperative. On the surface, of course, there seems to be something wrong, because she is manifesting that she did not capture the illocutionary force of my request. However, to the extent that I get her joke, presumably we cannot hold that she has not been IC-cooperative.

play a determining role in the generation of specific effects related to the degree of plausibility hearers may ascribe to the content conveyed by the speaker.[23]

Now, when it comes to explain specific mechanisms of meaning derivation, pragmatic accounts diverge on a number of issues, notably on the scientific adequacy of Grice's CP. Neo-Griceans take it on board while post-Griceans do not. According to the latter, and this is something which can also be concluded from the preceding discussion on the difference in nature between the three kinds of cooperation, there is much room left to question the properly cooperative status of both CC and IC. Common sense seems to play a role in this questioning, since we intuitively grasp a qualitative difference between cooperating to be understood and cooperating, say, to build a tree hut. The outcome of such a discussion would depend on what we exactly mean by 'cooperation'. Relevance theorists – who seem to favour the second acceptation of 'cooperation' (cooperation towards extra-linguistic goals) – doubt the very idea that the mechanisms involved in communication hinge on cooperativeness requirements. Cognitive approaches to communication such as RT not only dispose of cooperation, in the first two senses evoked earlier, but also tend to consider PC to fall outside the scope of an account of communication. Perlocution is accordingly seen either as a mere consequence of communication or as a purpose leading people to engage in meaningful communication; but this, according to RT, is only peripheral to the mechanisms underlying communicative exchanges and which make them possible. As for CC and IC, Relevance theorists dismiss their cooperative status, which leads them to consider that the way our cognitive machinery works does not involve a CP. In short, they assume that there is no need to postulate cooperation in order to explain how communication works.

I.1.7. POST-GRICEANS: RELEVANCE THEORY

While neo-Griceans make cooperation a pivotal point of the theory, post-Griceans disagree with the necessity of a CP. Relevance Theorists in particular usually consider linguistic cooperation to be a case of coordination at best (see Allott 2007: 14-16), when they do not completely dismiss cooperation from their explanatory account (see Wilson 1994). In what follows I summarise these views against the necessity of a CP and try to link CC and IC to the very fact of being relevant.

[23] The question arising here is to assess whether this functions as an overarching principle or as a contextual assumption among others which enters the process of meaning derivation. See §I.1.7. and §I.1.8. below for a discussion.

I.1.7.1. Why the CP does not explain comprehension and communication

In her (1994) paper, Wilson recalls that for Grice, "every utterance, every conversation, has an accepted purpose or direction whose identification plays a crucial role in comprehension" (1994: 35). She then identifies at least two problems with Grice's CP: if the principle stipulates that communication must bear an "accepted purpose or direction", how do language users identify the latter, and, moreover, how, if identified, does it help with comprehension? In other words, she questions whether identifying the direction in which the exchange is going (e.g., identifying the perlocutionary purpose of the exchange) is helpful in the derivation of the actual intended meaning. To take her example where a speaker says of his new tennis partner that

(11) He has much in common with John McEnroe

how would knowing that the purpose of the exchange is to characterise the speaker's new tennis partner help to infer that he is temperamental, instead of inferring that he, like John McEnroe, has a brother who also plays tennis, that he is also a rock musician, that he is North-American, that he is left-handed, and so on? Why would knowing that conversational participants are cooperating tell the hearer which particular representation he must pick, among all possible representations concerning his new tennis partner and John McEnroe the speaker is actually intending to communicate? Of course, for people who are familiar with the world of tennis in the 1980s, McEnroe's temper is a very salient assumption; nevertheless, there is *a priori* no reason to straightforwardly correlate salience, taken as such, and cooperation. Even if we consider that the speaker, by assuming the particular salience of McEnroe's temper, can be said to cooperate, in that she expects the hearer to pick up the salient assumption, next to nothing may formally allow us to justify that relying on salience is a trace of cooperation, and this by no means explains why or how the hearer should get the appropriate interpretation. Going back to (11), there is in principle no reason to assume that knowing that we are talking about how the speaker's new tennis partner is like will guide the hearer's selection of the relevant assumption (among all possible assumptions concerning the properties one is likely to ascribe to John McEnroe); this is rather linked to properties which can be ascertained about John McEnroe, and which are independent of any communicative endeavour. Thus, one major objection to the CP casts doubt on whether it is useful and necessary in processes of comprehension or not. This attacks the core assumption in Grice's model according to which the CP is required to manage meaning, i.e., that it is drawing on the CP that people may understand an utterance. As Wilson and Sperber note, "it is not at all clear that it is *because* conversation is a co-operative venture that utterances are interpreted in the way they are" (Wilson & Sperber 1981: 175).

An even more serious theoretical problem with Grice's CP, according to Wilson, is that assuming that the identification of the shared purpose plays a role in comprehension does not tell us much

about how this identification takes place. In fact, she considers the claim according to which the identification of the shared purpose of the exchange is achieved via the CP to be circular, since such a claim would translate into something like 'language users identify the shared purpose of the exchange through a principle that ensures that they do share a purpose in the exchange'. In these terms, in order "to identify the purpose of an utterance by use of the Co-operative Principle, one would already have to know it" (Wilson 1994: 36). Unsurprisingly, she concludes that "there is no Co-operative Principle" (*ibid*.).

In fact, Relevance Theory assumes a much weaker degree of cooperation in communication than Grice does in so far as it treats cooperation as a contextual assumption, not as an overarching principle of communication that explains why we derive the meanings we derive. All in all, Wilson's claim amounts to considering that cooperation happens to be a property of some communicative exchanges, and not some kind of rational rule without which meanings cannot be exchanged. From this perspective, communicators do not need to share a mutual direction in order to be able to communicate.

This line of argument was particularly tackling the issue of perlocutionary cooperation; we will nevertheless see that Relevance theorists also consider that what I refer to as CC and IC are not needed to explain how communication works.

I.1.7.2. Allott: coordination vs. cooperation

Wilson's objections were aimed at questioning the relevance of PC in an account of communication. But there is also some work within the framework of Relevance Theory that files an attack on CC and IC. The dismissal of the necessity of a CP that would underlie the calculation of meaning is expressed in Allott's distinction between coordination and cooperation. In his doctoral thesis (Allott 2007), he argues that what I call CC and IC are not properly cooperative aspects of communication, but coordinative ones:

> "As far as communication is concerned, wanting to be understood does not entail wanting to contribute to a joint undertaking. [...] While I agree that speakers generally intend their utterances to be interpretable, I regard that as coordination arising from general rationality considerations, not cooperation. It is rarely worth saying something to someone else unless you intend to be understood. [...] So making an utterance with the aim or intention of being understood is not cooperation, but simply the rational interest in one's action succeeding (as argued in Sperber & Wilson, 1986, pp. 161–162; Sperber & Wilson, 1995, pp. 267–268). For a communicative action to succeed as such, it must be interpretable." (Allott 2007: 15-16)

Allott considers that a speaker's willingness to convey a meaning and a hearer's willingness to interpret a speaker's utterance are two independent intentions which happen to converge in the broader frame of communication, and that they do not constitute a cooperative effort, since for him the requirement for successful communication seems to be interpretability: "for a communicative action to succeed as such, it must be interpretable" (Allott 2007: 16). He takes these two actions to be rational underpinnings of communicative behaviour, on each of the participant's behalf, but not a cooperative effort. I would venture that the reason behind this may have something to do with the consensual conception of cooperation as a somewhat conscious and (possibly) planned endeavour: it indeed seems on the face of it that working together to build a tree hut requires much more cooperation than merely getting a message across. The question is whether it is sound, in terms of cooperation, to consider that speaker and hearer cooperate in making sure that meaning comes across. Allott seems to take the automatic nature of comprehension to leave no room for considerations about speakers and hearers actually sharing a goal; in other words, he does not take the goals of wanting to be understood and willing to understand to merge into what could be taken as a shared goal of conversational participants: "The speaker and hearer are engaged in different activities, with different aims, and this is not cooperation in the Gricean sense, which requires that the talk-exchange have a shared aim or purpose" (Allott 2007: 15).

But let us examine Allott's claim more closely: "For a communicative action to succeed as such, it must be interpretable" (Allott 2007: 16). I would add "and interpreted" right next to "it must be interpretable", for it is questionable to consider the completeness and success of a communicative act from only one perspective, which, if we strictly focus on Allott's preceding quote, seems to be the speaker's. If no one is there to actually interpret an utterance, we can hardly say that communication has been successful, let alone that it has taken place, even if the *attempt* to communicate has. From the perspective of conversation, the "maximally effective *exchange* of information" Grice speaks of (Grice 1989: 28, my italics), communication cannot be described as a specific type of interaction between two rational agents without taking into account both of their behaviours, that is, production *and* reception. Indeed, one can argue that successful and complete communication supposes not only that some content is encoded in an utterance in a particular way, but also that that content is retrieved by a hearer. Otherwise we are only describing one end of the process, or its necessary conditions, without considering the phenomenon in its entirety.

This echoes something Sperber and Wilson evoke in their book on relevance; when they equate communication with the recognition of intentions, they indeed consider that "communication is successful not when hearers recognise the linguistic meaning of the utterance, but when they

infer the speaker's 'meaning' from it" (Sperber & Wilson 1995: 23), which presupposes that the success of communication depends on whether there is a hearer that will pick up the intended meaning: "communication is achieved by the communicator providing evidence of her intentions *and* the audience inferring her intentions from the evidence" (Sperber & Wilson 1995: 24, my italics).

While I do not deny the crucial role of intelligibility and interpretability in communication, together with their being expected by conversational participants (which I have attempted to characterise through the notions of CC and IC), I take the actual interpretation of the stimulus to be part of the picture as a condition without which successful and complete communication would not take place. Imagine I go to the Sahara desert, that I am alone, and that I shout, out of despair, "I'm thirsty!" Can we consider the scenario as successful communication? Allott probably would say so, in that what counts is my utterance of an intelligible content. However, the goal I may be pursuing by making my utterance, which is arguably, in a minimal way, that someone understands what I'm saying, will not be fulfilled unless someone is actually there to hear and interpret my utterance. One could argue about this example that my utterance of "I'm thirsty" in the Sahara desert, to the extent that I am aware that no one is there to pick it up, is by nature not aimed at conveying a content to someone, but rather at expressing something else, say my frustration, even if it is to myself, regardless of the fact that I know I have no addressee. But then this would not be communication proper, unless we consider that I am my own addressee, and that I want to communicate something to myself.

Another example could be one where I scream for help in a situation where I need someone's assistance, for instance in a car crash on a Sunday morning, in a deserted street. Here, even if my request for help is formulated with the extra-linguistic goal of having someone assist me, we cannot say that communication will successfully take place unless someone understands my message. In other words, although my utterance may very well display all the required features of rational communication, this will not really matter, and certainly not ensure successful communication *until someone interprets it*. A simple answer would be to say that since cooperation for communicative purposes by definition requires at least two conversational participants, my example is not a complete instance of communication, but rather a mere attempt to communicate, that will be successful the moment an addressee interprets my message. Still, in this case, this is precisely my intention when I scream for help.

The question arises once again: if coordination, as postulated by Allott, does justice to communication, can we still use the notions of CC and IC as relevant categories to describe and explain communication, and, what is more, as relevant principles of communication? CC is a crucial requirement of communication, given that it allows distinguishing meaningful content

from plain gibberish, as noted by Chilton (1987). However, this suggests that it is an inherent feature of communication more than a principle of communication which plays a role in people's communicative behaviour, either on the production or the reception end of the process. CC, as Wilson's critique turns out to show, is not a principle that will guide language users in the derivation of meaning because nothing in its formulation concerns the way meanings are actually managed in conversation; rather, it is a necessary condition which ensures that meaningful exchanges will take place, ensuring both production and reception. Therefore, CC is about putting our communicative rationality to use by indicating readiness to engage in communicative exchanges, and as such it a definitional feature of communication, though it can hardly explain how it is that language users end up entertaining specific representations (and not others) through communication. In fact, and as ventured earlier, I would be inclined to consider that CC is an aspect of human rationality in communication; in this respect, my position is not so different from Allott's, with the difference that I would still speak of cooperation rather than coordination, since i) the success of the utterance depends on whether a hearer picks it up or not and ii) I treat communication from the perspective of the *exchange* of information.

This echoes a very important point made earlier in contemporary philosophy of language by Strawson. In his (1964) paper, he accounts for communication by considering the uptake of speech acts to be one of its necessary conditions: a speaker "will not have secured understanding of the illocutionary force of his utterance, he will not have performed the act of communication he sets out to perform, unless his complex intention is grasped" (Strawson 1964: 450). It is quite clear, in the second half of the quote, that Strawson considers that communication requires the addressee to recognise whatever intention the speaker wanted to communicate through the formulation of linguistic information. This supports the claim according to which communication as a whole is more cooperative than coordinative.

The reason for this is the following: if we take into account the *success* of the communicative act in addition to its necessary conditions, we end up with the necessity of postulating the intervention of a hearer that will actually interpret the utterance. In terms of goals, and once we acknowledge the necessity of the hearer, therefore, it does make sense to speak of a joint endeavour: both speaker and hearer need to be willing to interact about meaningful contents if communication is to take place.

Sperber and Wilson also point to this very idea: "we ourselves have stressed that interlocutors always share at least one common goal, that of understanding and being understood" (Sperber & Wilson 1995: 268). Goal-sharing *is* cooperation, and this is exactly my point in the characterisation of CC.

In disposing of the CP, Sperber & Wilson actually dispose only of PC (since lying and deceiving remain plausible communicative phenomena), but CC and IC can still be salvaged in their account, though in other terms than those Grice used, as suggested by the preceding quote: there is indeed a shared goal in communication, which can translate into cooperation. However, the power Sperber & Wilson grant to cooperation is much more limited, compared neo-Griceans. It is not a *principle* of communication in RT, but a contextual assumption a hearer may form about the speaker's intentions and mental states. As such, it is not an overarching norm by which meaning is made possible. If we are looking for requirements of communication, under this view, we need to address natural cognitive mechanisms of information-processing and inference. We will see that it can still make sense to speak of cooperation, but not in a Gricean sense, that is, not considering cooperation as an *explanatory* principle.

I.1.8. THE DESCRIPTIVE VALUE OF COOPERATION

When Chilton attempts to define communication *simpliciter*, he postulates that there is "strong pressure to assume relevance, that is to assume that the speaker observes the CP, that he or she is 'co-operating'" (Chilton 1987: 224). Even if he remains on the Gricean track by keeping the CP, his mention of relevance echoes some core claims of Relevance Theory and at the same time my own characterisation of CC as a necessary condition of rational communication. Indeed, when a speaker produces an ostensive-inferential stimulus, she thereby communicates to her addressee that this stimulus is worth processing; it accordingly carries a presumption of optimal relevance (this is the whole point of recognising the speaker's *communicative intention*). If we construe this natural behaviour from a goal-driven perspective, we can easily translate it into cooperative behaviour, because being relevant (in Sperber & Wilson's sense) is a realisation of both the speaker's readiness to communicate (i.e., CC), and her readiness to communicate meaningful information in a standard, rational way (i.e., IC), namely her readiness to produce a stimulus which is rational for the purposes of communication.

For communication to succeed, the speaker's communicative intention must be recognised by the hearer: in light of the tripartition of cooperation evoked in the preceding sections, the speaker's communicative intention supposes that the she is being cooperative on the CC and IC levels, and, in some cases, that she is also calling for PC on behalf of the hearer. Even if there are doubts to consider that cooperation (in any of the three senses I have identified) is a proper *principle* which guides the hearer's processing of information, it nonetheless remains that a speaker's communicative behaviour *is* cooperative, because it satisfies linguistic and sometimes extra-linguistic goals which can to a large extent be considered as shared by conversational

participants. And the moment there is goal sharing, in my opinion the possibility of cooperating can be envisaged.

Thus, cooperation, as could be hosted in a relevance-theoretic framework, neither has the causative nor the explanatory power a Gricean approach would advocate. Sperber and Wilson actually do not agree with the idea of equating the production of presumably relevant contents with cooperative behaviour; they state that "it would be a mistake to read the presumption of optimal relevance […] as describing a goal that rational communicators should achieve" (Sperber & Wilson 1995: 271). What they want to make clear is that speakers are not relevant *in order to* cooperate; however, this does not deny the possibility of considering that, in being relevant, they may be *described as* cooperating. The difference is subtle, yet important, since it allows RT to dismiss the necessity of a CP, on the grounds that *cooperation*, for the analysis, only bears a *descriptive role* and not a normative one:

> "Unlike Grice's maxims, neither the principle nor the presumption of relevance is presented as a goal to be pursued or a rule to be followed by the communicator. The (Second) Principle of Relevance is a descriptive (as opposed to normative) claim about the content of a given act of ostensive communication. It claims that part of that content is a presumption that his very act of communication is relevant to the addressee." (Sperber & Wilson 1995: 271.)

So while Sperber & Wilson would probably not disagree with the idea that we can describe the fact of being relevant as being cooperative, they reject the view according to which we are relevant *because* we want to cooperate, in other words, that we are relevant by virtue of a principle of cooperation that governs communicative exchanges. This is the difference between an RT-inspired descriptive status of cooperation, and Grice's normative status of cooperation.

Indeed, as soon as we consider the CP to be a norm, questions as to its legitimacy and origin emerge. In case we acknowledge the normative nature of the CP and the maxims, as noted by Wilson, we could ask the following:

> "Is co-operation essential in communication? Do speakers really aim at truthfulness, informativeness, relevance and clarity? […] Where do the Co-operative Principle and maxims come from? Are they universal? If so, are they innate? Are they culture-specific? If so, do they vary, and how are they acquired?" (Wilson 1994: 35)

All these questions prove rather thorny to answer, but in my view they can be removed from the account by establishing a clear distinction between the *explanation* of how communication works and the *description* of what communicators actually do in conversation. RT's solution to the problem is to assume a much reduced degree of cooperation in communication, and to

devoid it of any causative power in processes of meaning derivation. To go into more detail, the aforementioned distinction between CC and IC on the one hand, and PC on the other, may come handy.

If we acknowledge that communication fulfils specific goals or satisfies particular purposes (be they informative, argumentative, social, and so forth), then any communicative attempt whose success relies on the production of a stimulus *and* on its recognition as a meaningful one (and on the mutual manifestness of intentions) can be described as a goal-oriented activity. Accordingly, and insofar as its success requires that both conversational participants act in an effortful and rational way, any communicative exchange may be described as cooperative. This becomes very apparent for CC and IC; if conversational participants do not display readiness to communicate (i.e. to encode contents, speak, listen, and understand), then communication cannot take place. So, by default, communication is cooperative, in this very restricted, rudimentary first sense, because conversational participants must make these efforts if they want to secure proper communicative exchanges. Note that this is quite different from postulating that it is because communicators aim at cooperating that they act and process information the way they do.

As far as meaning is concerned, communication fails when the hearer fails to recover the intended content; symmetrically, when communication is successful, it can be said that the behaviour of conversational participants has been cooperative in view of the fulfilment of a maximally effective exchange of information, thus ensuring that the intended meaning was picked up appropriately. Again, this is quite different from assuming that their behaviour is governed by cooperative requirements. For that matter, any case of misunderstanding will only reveal that communication was not successful,[24] but this does not necessarily mean that it is because communicators were not able to IC-cooperate that communication failed.

It nonetheless remains that it still makes sense to *describe* speaker and hearer behaviour in communication as cooperative, both in CC and IC terms. That is not to say that their contributions are designed to cooperate or to observe a CP, nor that it is because they cooperate that they understand each other; rather, a cognitive pragmatic account would stress that it is the cognitive abilities of both participants that make cooperative behaviour possible.

The case of PC is perhaps simpler, because it scopes over extra-linguistic aspects of communication. In making an utterance, a speaker may or may not be pursuing some goal which is independent, in nature, from the actual success of the conversational exchange. That is to say

[24] Nevertheless, note that in misunderstandings CC does succeed; it is not because I did not understand the speaker that I did not try to do so. So misunderstandings are actually an indication of IC gone wrong, even if CC was still observed.

that communication does not require such a goal to be shared by both communicators. In Relevance-theoretic terms, as exposed in the preceding section, the basic requirement of communication is that conversational participants end up sharing a similar representation.

It appears quite difficult to disentangle the complexity of the issue; on the one hand, postulating that CC and IC are necessary conditions for the success of communication resonates with the intuitive idea that communication requires cooperation. But on the other hand, assuming that PC is a requirement of communication is misleading, given that communication may succeed without any extra-linguistic goal being instantiated by any of the communicators. The trichotomy I proposed is thus helpful only to the extent that it allows precisely characterising the role of cooperation in communication, according to its nature and scope.

When Wilson (1994) argues against the necessity of cooperation, she does so on two counts: first of all, she argues against PC (by noting that we do not need to know the purpose of an exchange to understand utterances produced therein), and second, she rejects the causative role a cooperative principle might have in an account of communication (and in this respect it implies that what I've called CC and IC are not principles of communication). Here is how she justifies this:

> "There is no Co-operative Principle, and hence no circularity in assuming that the purpose of an utterance can be identified, where necessary, as part of the comprehension process. To the extent that the purpose of an utterance does contribute to comprehension, it is identifiable as a contextual assumption like any other, via the criterion of consistency with the principle of relevance. In this framework, there is room for a notion of purpose, *but the real burden of explanation lies elsewhere*." (Wilson 1994: 37, my italics)

The identification of a purpose and the willingness to behave according to that purpose (which makes the exchange cooperative) do not explain why it is that communication is most of the time successful. However, there is nothing wrong in describing an exchange as an interaction that turns out to be cooperative, as long as we do not assume a causative link between cooperation and comprehension. And when conversational and situational circumstances make it relevant, the identification of cooperation may play a role in the derivation of meaning, but, as Wilson suggests, this is merely a "contextual assumption like any other" (*ibid*.), and not the reason why comprehension takes place.

I have shown that, though they are not granted explanatory force, it still makes sense, from the philosophy of language point of view, to describe communicative exchanges as communicatively cooperative interactions (in the CC and IC acceptations of cooperation). Wilson states that unlike Grice's account, "Relevancy Theory is not a rule-based or maxim-based system. […] Relevance is

fundamental to communication not because speakers obey a maxim of relevance, but because relevance is fundamental to cognition" (Wilson 1994: 36). Furthermore, she assumes that relevance is "an exceptionless generalisation about what happens when someone is addressed", and that "it makes no sense to claim that the principle of relevance can be overtly violated to create an implicature" (Wilson 1994: 36). Therefore, there are good grounds to consider relevance to bear a fundamental role in what I have referred to as rationality for communication, and equally good grounds to consider CC and IC as a by-product of relevance, for that matter. Being relevant yields cooperative behaviour: the speaker makes efforts to formulate an intelligible and interpretable utterance while signalling to her addressee that she intends to communicate something, and in parallel the hearer makes efforts to interpret the stimulus; in doing so, he solicits his natural readiness to communicate.

A last observation about a tentative relationship between relevance and cooperation has to do with the effort/effect dynamics Relevance Theory builds on. The basic idea is that cognitive mechanisms operate under economic constraints: the less effort a cognitive operation involves, the more advantageous its outcome is; in parallel, the more effects a cognitive operation yields, the more advantageous its outcome is. We may consider that cooperative behaviour is indeed a way to maximise effects while minimising efforts, in so far as it allows yielding benefits which would not be attained, or attained at a greater expense, without cooperators working towards the same goal (this is why cooperation is necessarily purposeful). This, in turn, is an indication that Relevance Theory makes use of, though restricting its scope to cognitive mechanisms, a dynamics which is similar to the one that underlies cooperative behaviour. The difference being that cooperation is not limited to cognitive considerations, since it can scope over behavioural phenomena. Yet, we have seen that CC and IC, as we have defined them, are rational aspects of communication, and that they can to some extent describe cognitive features of communicative exchanges (for they involve a readiness to communicate and the management of specific meanings in conversation). In principle, therefore, there is room to defend the idea that some kinds of cooperation may still be relevant to a cognitive pragmatic account of communication such as Relevance Theory, albeit in a descriptive fashion.

I.1.9. CONCLUSION

I have tried to show in this section that when we envisage communication as directed towards the fulfilment of specific purposes, it makes sense to describe it as a cooperative endeavour, since the attainment of certain goals can reasonably be taken to act as powerful purposes for communication. I have also tried to defend the idea that this holds for a perspective in which successful communication is construed as an interaction involving effortful cognitive behaviour

from at least two participants. These can cooperate i) towards the possibility of communication (CC), ii) in the way actual exchanges of meaningful contents unfold (IC), and iii) towards the fulfilment of extra-linguistic goals (PC). Now, this does not mean that the identification of a purpose must play a decisive role in an explanatory account of meaning. I have exposed to this effect Wilson's arguments, which convincingly show that the explanation of communicative phenomena can do without cooperation.

However, I have also argued that it still makes sense to speak of cooperation on at least two counts: first, what Relevance Theory calls 'being relevant' (construed as a feature of cognition) may turn out to be *described* as CC and IC within the frame of the communicative interaction; such a distinction could in turn enrich the descriptive account of communication by developing the rationale for considering these two kinds of cooperation as rational features of communication. Second, although it is not necessary for the success of communication, PC might help us discriminating specific instances of communication in that it can determine the outcome of certain exchanges (remember Lumsden's (2008) example where an utterance, whether formulated by a car dealer or a close friend, might elicit two different responses).

To sum up, let me list three conclusions this section has reached; these are aspects of communication that make explicit the relationships between the three types of cooperation envisaged and relevance:

- CC can be expressed as a natural readiness to communicate, a rational disposition to produce and interpret meaningful ostensive stimuli. It is at play the minute the hearer recognises the speaker's communicative intention embedded in the stimulus; furthermore, it expresses the recognition, by the hearer, that the speaker is formulating a relevant utterance which is worth processing. On the speaker's side, it amounts to the mere production of a meaningful stimulus destined to be picked up by an addressee. Overall CC is definitional of communicative exchanges, for its absence would prevent the goal of exchanging information from ever being fulfilled.
- IC describes the way speakers and hearers manage meaningful contents; it is achieved when the hearer has worked out speaker meaning; it is the recognition of the *actual* contextual relevance of the speaker's utterance by the hearer. As such, it supposes that the speaker encodes her informative intention in a way which allows the hearer to work it out, and thus IC expresses the idea that speaker and hearer both make use of the same cognitive mechanisms in order for meaning to be successfully exchanged.
- PC refers to the idea that people cooperate, through communication, towards goals which are beyond the communicative exchange. It is achieved beyond the exchange of

relevant information, is unnecessary to the success of a communicative exchange, and supposes further confirmation (verbal or plainly behavioural) on behalf of the hearer.

In so far as it bears descriptive value, I will continue to use this tripartition as I proceed to spelling out what it means to be non-cooperative in communication in the next section.

I.2. NON-COOPERATION

I.2.0. INTRODUCTION

In the previous section, I have discussed, from a pragmatic perspective, the nature and implications of cooperation with regard to communication, arguing for a tripartite conception of cooperation mostly relying on the type of goals one may want to pursue in and through communicative exchanges. Now that the descriptive foundations of cooperation have been laid, I can start assessing non-cooperation in communication. The aim of the following section is thus to establish to which extent Gricean theory (including the elaborated conception of cooperation I proposed) may serve in the description and explanation of manipulative communication. For the time being I shall use the terms 'manipulative' and 'deceptive' somewhat indiscriminately. I will though, in the next sub-chapter (see §I.3.3.), briefly discuss their semantic differences.

My first task will be to discuss what it could exactly have meant to Grice to be uncooperative in a communicative exchange, by first reviewing his developments of conversational moves which fail to fulfil the maxims. I will be led to postulate that there are different types of non-cooperation (basing my discussion on the tripartition I operated above between CC, IC and PC), and that Grice's account of meaning construction (by observance of the CP and the conversational maxims) can achieve only a partial model of non-cooperative communication. I will argue that this is in fact not surprising, given that Grice's model is not meant to address phenomena of deception, but rather instances of non-cooperation which are still, at least to a certain extent I will try to characterise in detail, cooperative cases of communication. Accordingly, I will be led to postulate covertness as a decisive criterion to characterise uncooperativeness, drawing on research within Information Manipulation Theory (cf. McCornack 1992); to this effect, I will assess what could have been Grice's own notion of conversational non-cooperation in order to establish that the latter is not to be put on a par with the notion of non-cooperation (leaning towards manipulation) I am concerned with in the present work.

Since my goal in this section is to give a tentative account of what non-cooperation would look like in the Gricean theory and to expose its limitations, I will follow that particular framework in the first part of the exposition, even if further along I intend to expose its limitations. Let me thus 'make as if' I actually adhered to all Gricean precepts for the time being.

I.2.1. GRICE ON NON-COOPERATION

I.2.1.1. Maxim violation and non-cooperation: preliminary remark

Though Grice was interested in accounting for how utterances may mean something different from their semantic content through observance of the Cooperative Principle and the operation of conversational maxims, most of his work on implicature is devoted to the discussion of cases where the latter are not observed. After having stated his Cooperative Principle (CP) and the four conversational maxims (Grice 1989: 26-30), he set out to examine four types of cases in which conversational participants fail to fulfil the maxims. His goal in doing so was to examine the "connexion between the Cooperative Principle and maxims, on the one hand, and conversational implicature on the other" (Grice 1989: 30), so as to provide a model which could be able to explain the underdeterminacy of semantic meaning.

Much of the debate in post- and neo-Gricean pragmatic literature dealing with the maxims and their relation to meaning originates in the list he draws of failures to fulfil the maxims (particularly cases 3 and 4 below), which I reproduce here:

(12) Types of failures to fulfil conversational maxims (following Grice 1989):

> "A participant in a talk exchange may fail to fulfil a maxim in various ways, which include the following:
> 1. He may quietly and unostentatiously *violate* a maxim; if so, in some cases he will be liable to mislead.
> 2. He may *opt out* from the operation both of the maxim and of the Cooperative Principle; he may say, indicate, or allow it to become plain that he is unwilling to cooperate in the way the maxim requires. He may say, for example, *I cannot say more*; *my lips are sealed*.
> 3. He may be faced with a *clash*: He may be unable, for example, to fulfil the first maxim of Quantity (Be as informative as is required) without violating the second maxim of Quality (Have adequate evidence for what you say).
> 4. He may *flout* a maxim; that is, he may blatantly fail to fulfil it. On the assumption that the speaker is able to fulfil the maxim and to do so without violating another maxim (because of a clash), is not opting out, and is not, in view of the blatancy of his performance, trying to mislead, the hearer is faced with a minor problem: How can his saying what he did say be reconciled with the supposition that he is observing the overall Cooperative Principle? This situation is one that characteristically gives rise to a conversational implicature; and when a conversational implicature is generated in this way, I shall say that a maxim is being *exploited*." (Grice 1989: 30, author's italics)

These four instances are characterised by Grice as non-observances of the conversational maxims, or, as he puts it, as failures to fulfil them. Were we to consider Grice's as a normative system, a possible viewpoint on this issue would be to consider that strictly speaking, i.e. as far as the observation of the maxims is concerned, these non-observances should be symptomatic of uncooperativeness. If we take the maxims as conversational standards – more as rule-like principles – ensuring cooperativeness, any such non-observance would therefore count as uncooperative.

However, such a characterisation does not stand basic criticism. For one thing, Grice does not consider maxim observance as a necessary condition for cooperativeness; his model indeed includes the possibility of failures to fulfil maxims as overall cooperative conversational moves (e.g. (13) and (14) below), which suggests that cooperation is not determined by proper observance of the maxims. Grice rather considers cooperation as a manifestation of communicative rationality. More specifically, in order to dispose of a potential argument which would assume that any contravention to the maxims results in non-cooperation, he distinguishes two levels of meaning at which maxims can be violated: the level of what is said, and the level of what is implicated: while discussing the category of examples involving maxim exploitation, he notes that "[…] though some maxim is violated at the level of what is said, the hearer is entitled to assume that that maxim, or at least the overall Cooperative Principle, is observed at the level of what is implicated." (Grice 1989: 33)

So while a speaker may be literally uncooperative on the level of what is said by failing to observe the requirements of one of the maxims in the explicit content of her utterance, she might turn out to be completely cooperative on the level of what is implicated: the operation of the maxims and the CP still obtains, albeit with regard, in some cases, to implicit contents. What should be noted in any case is that cooperation, in Grice's framework, is characteristic of successful meaning exchanges.

By Grice's standards, it would thus be a mistake to consider cooperativeness at face value (in its technical acceptation), i.e., as the sole expression of strict maxim observance.[25] Rather, cooperativeness is a basic assumption language users hold – arguably tacitly – to obtain in all conversational exchanges, particularly in the IC sense of cooperation. In other words, whether a

[25] Some of the criticism Grice faced was accordingly ill-formulated according to Davies (2007): some researchers have (wrongly) assumed that his model was overly ideal, grounding their criticism on real data showing non-observance of the conversational standards he set. Obviously, Grice was well aware that cooperation in a very restricted and technical sense was not the type of cooperation he took to regulate conversational exchanges and deployed much of his efforts to examine cases in which maxim violation ensured a maximally effective, thus rational, exchange of meaning.

speaker is being IC-cooperative cannot be established by looking at whether she is observing the maxims or not, since maxim observance and maxim non-observance may both ensure a successful exchange of meaning.[26] When conversational participants observe the maxims they are usually cooperative; however, they can also flout a maxim in order to convey implicatures which they intend their addressee to derive. Thus, it is predicted by Grice's model that IC may result from both observance and non-observance of the maxims, the only requirement of non-observance being that it must be overt enough for the addressee to recognise it. IC is therefore not about strict observance of the maxims; it is rather about making sure that the addressee recognises non-observances when they occur, and that he assumes that maxim observance obtains, unless prompted to consider otherwise. A natural conclusion to be drawn from these remarks is the following: strict maxim observance is no definite criterion for ascertaining IC, for IC obtains both when one observes of fails to observe the requirements of conversational maxims, as long as this is overt.

This last consideration finds straightforward illustration in non-literal usages. Were we to define IC in Gricean terms as a scrupulous and rigid observance of the maxims, some extremely frequent and very natural possibilities of figurative speech and implicit communication would be ruled out, for these usually involve blatant failures to observe the conversational maxims.

Take for example the case of metaphor, which Grice (Grice 1989: 34) discusses in light of its strict non-observance of the maxim of quality. When I utter something like

(13) Laszlo is a lion.

I neither reasonably expect my audience to infer that I believe that the individual I refer to as Laszlo is actually of the species of large wildcats usually found on the African continent, nor do I expect my audience to take me as being committed to this state of affairs. Rather, my point is to describe Laszlo as a cruel, courageous or proud individual (and so on, according to the many properties one can attribute to lions), depending on the context. In other words, the fact that truth-conditions are not met in metaphorical usages, although technically amounting to a violation of Quality, does not necessarily entail that the speaker is being uncooperative.[27]

[26] Recall that I defined IC as informative cooperation, i.e., the level of cooperation where conversational participants produce and interpret meaningful contents.

[27] But this does not mean that metaphors may nevertheless not be used for manipulative purposes. See for instance Chilton (2005) for an illustration of this idea based on an analysis of Hitler's *Mein Kampf*.

Similarly, I cannot expect my audience to infer that I actually believe that I am the Queen of England when I utter (14) in response to my friend Laszlo's unrealistic assertion according to which he will finish his PhD within the next two months, as I perfectly know he will not make it.

(14) And I am the Queen of England.

(14') I do not believe that you will be done in two months' time.

My utterance, in this case, would rather communicate something like (14'), and does obviously not commit me to actually believing that if Laszlo did finish it within the following two months I would honestly consider myself to be the Queen of England. Again, in this case the maxim of quality is not observed, but it need not automatically involve my being IC-uncooperative in the exchange.

Most of Grice's discussion on maxim non-observance unfolds along this spirit, and according to his account, even if such examples are truly failures to fulfil the maxims they cannot be equated with breaches of the CP in the IC sense. The justification for this is that conversational speaker meaning usually manages to go through (which makes communication successful) notwithstanding maxim non-observance, because audiences are expected to spot and understand these without this automatically resulting in inferring a breach in the CP. In other words, these failures to observe the maxims are not IC-uncooperative since they are manifest and intended to be so.[28]

What the preliminary examples of metaphor and irony demonstrate is that *strict non-observance of conversational maxims is not necessarily to be put on a par with a violation of the CP*, i.e. with non-cooperative conversational conduct, as they still ensure meaningful exchanges (i.e. they abide by IC). In what follows, I will explore more in detail the issue of non-cooperation and its relation to the operation of maxims by examining Grice's list of ways in which speakers may fail to fulfil the maxims (reproduced in (12) above).

I.2.1.2. Ways of failing to fulfil conversational maxims

Grice's motivation to characterise maxim non-observance was not to examine in which ways one can be uncooperative in a conversational exchange; we have seen earlier that he was after a characterisation of implicature, taken as a specific rational manifestation of the way we communicate, and not so much after a proper and detailed model of cooperation. In fact, he was more interested by a way to account for the fact that non-observance of maxims can many times

[28] This suggests that contraventions to IC do not necessarily rely on failure to observe the maxims. See section on IMT for a detailed discussion (§I.2.3.)

be straightforwardly cooperative. If we want to focus properly on uncooperativeness from a Gricean perspective, we will thus need to develop his list of different ways to fulfil maxims in light of considerations specifically related to uncooperativeness, so as to assess, in his terms, what exactly it means for a conversational participant to be uncooperative.

I.2.1.2.1. Unostentatiously violating a maxim

This is the first of the four ways in which Grice envisages non-fulfilment of maxims. Little is said about it; all we can find in Grice (1989) is that such violations correspond to cases where a speaker "quietly and unostentatiously *violate*[s] a maxim"; and that in some of these cases, Grice proceeds, the speaker "will be liable to mislead" (Grice 1989: 30).

This is precisely the type of non-fulfilment I am after as far as the topic of the present thesis is concerned, since it corresponds to *covert* uncooperative cases of communication (which Grice calls misleading, and which I would take to fall under the broader category of manipulative communication). These typically involve some kind of asymmetry, for only the speaker is aware that maxims have been violated. That is, the speaker manifests standard communicative behaviour (i.e., she does not communicate that the requirements of conversational exchanges are not being met), but knows that she is deliberately failing to fulfil at least one maxim without making the failure public, i.e., without intending the violation to be recognised as intentional.

It is probably not a coincidence that Grice did not elaborate more on unostentatious maxim violation, arguably because the set of phenomena belonging to this category is different from instances of non-cooperation such as those we will expose next: maxim violation is considered to be *covert* in this first category of maxim non-fulfilment, while all other failures to fulfil the maxims are *overt*. Furthermore, Martinich (1984: 33-34) points out that Grice himself seems to have had trouble with this category, as can be seen in the example illustrating a purported violation, since the latter (illustrating a social gaffe, cf. Grice 1989: 35) is about a *blatant*, thus ostentatious, violation. Maxim violations are unostentatious by definition and consequently they cannot be blatant, that is, they should not be recognised by the hearer (which is something that does happen in Grice's aforementioned example).

Taking these elements into account, it straightforwardly appears that Grice's account of implicature and cooperative communication is somewhat unsuited to account for phenomena such as misleading, which must remain covert to be successful. This is because this particular type of non-fulfilment of the maxims (i.e. unostentatious maxim violation) seems to fall outside of the scope of phenomena his model was originally meant to capture. In addition, it also highlights that the crucial factor in the success of unostentatious maxim violation lies within its covert nature, and not in its contravention of communicative standards.

An additional argument can support the claim according to which covert maxim violation can hardly be accommodated within Gricean theory, and builds on a simple observation of what happens, in terms of communicative behaviour (especially on the reception end), when a speaker successfully misleads her audience. If we consider such instances of communication from the perspective of the hearer, it stands to reason that communication has unfolded in a completely standard way. If the latter is not aware of any violation – nor of any breach of cooperation for that matter – it cannot be argued that the way he processed the information in the calculation of meaning differs from that of standard exchanges; therefore, in principle there is no question of any putative difference between the *processing* of misleading and of non-misleading utterances. That is, of course, from the perspective of the hearer.

In view of adding detail to this exposition in light of the tripartition I operated between CC, IC and PC, we must first remember that whether a speaker unostentatiously violates a maxim makes absolutely no difference for the cooperative and trusting hearer, as the latter is unaware of the violation. As noted above, attempts to mislead are only successful when hearers do not notice them; accordingly, successfully misled hearers assume CC (since they are engaged in a communicative exchange), IC (in that they cooperate in the transmission of meaning and that they take the speaker to cooperate as well by taking her to endorse ostensive non-observances of the maxims when they occur), and most of the times even PC (unless they have reasons to doubt the speaker's perlocutionary cooperativeness).

This is by the way a supplementary indication that a very clear separation has to be made between ostentatious and unostentatious operation of the maxims; what is more, I take this to strongly support the claim according to which the Gricean theory of meaning would face great challenges were we to apply it in a tentative account of misleading and manipulative instances of communication.[29]

To sum up, we could say that unostentatious maxim violation is a category which seems to be at odds with Grice's theory on three counts: i) the phenomena belonging to this category are essentially distinct from the phenomena belonging to the other categories of failures to fulfil the maxims in that the former are *covert*, ii) Grice himself seems to have had trouble in characterising the category, as can be seen from the problematic example he gives, and iii) it appears that successful processing of misleading statements is completely standard, and

[29] In all fairness, it must nevertheless be acknowledged that Grice did not intend his theory to account for such phenomena; it would thus be probably a mistake, and certainly unfair, to criticise his work on these grounds, if only because he did not himself intend it to capture them. What can be done, however, is an assessment of how his theory accommodates such instances of communication. This is the main goal of the present section.

consequently, from the perspective of unwittingly misled hearers, whether there has been a covert violation makes absolutely no difference, for they are unaware of it and engage in meaning calculation along the standard procedure.

I shall return later on to this category in order to discuss in more detail an attempt to accommodate a Gricean framework (namely Information Manipulation Theory, McCornack (1992)) with deception, and will also take a few examples where unostentatious maxim violation is at stake.

I.2.1.2.2. Opting out

The second case in which a speaker may fail to fulfil the maxims is when she opts out "from the operation both of the maxim and of the Cooperative Principle" (Grice 1989: 30). From this formulation it appears that one can either opt out of the CP (i.e. refuse to make one's contribution such as is required by the accepted direction of the exchange) or of the conversational maxims (and also possibly of both at the same time). I will examine both cases in detail in order to assess to what extent opting out turns out to be (un)cooperative in communicative exchanges.

I.2.1.2.2.1. Opting out of the operation of conversational maxims

Opting out from the operation of the conversational maxims obtains when the speaker has communicated, either explicitly or implicitly, that she was unwilling or unable to observe the requirements of the maxim. In this sense, opting out denotes the *result* of a speaker's decision and public announcement not to comply with communicative standards.

Martinich (1984: 36) gives a series of helpful examples illustrating how speakers may opt out from the operation of all four conversational maxims. His short exposition of these is reproduced below:

a) Opting out of Quantity:

   By uttering "I don't remember, dear", after being asked for specific information, a speaker is failing to comply with the maxim of quantity because she is not able to provide the required information. As she is faced with a clash (between Quantity and Quality, i.e. between providing the right amount or information vs. providing true and reliable information), she will opt out of Quantity in order not to infringe Quality.

b) Opting out of Manner:

"Replying to the request, 'Explain Einstein's theory of special relativity to me without bringing in all those obscure technicalities', one might well say 'What you ask is impossible'" (Martinich 1984: 36). Here the speaker uttering the reply is communicating that she cannot meet the request because explaining Einstein's theory requires jargon. To the uneducated ear, such jargon is necessarily technically obscure, but not to the scientific mind, and thus a jargon-free scientific explanation is inconceivable. In other words, the speaker is conveying the idea that non-observance of Manner is unavoidable, and thus manifesting her opting out of the operation of that maxim.

c) Opting out of Quality:

Martinich gives the example of pretence as a clear case of a speaker opting out of Quality. To take his example, if I say "let's pretend that I am the Pope and you are the general secretary of the Polish Communist Party" (Martinich 1984: 36), I am opting out of Quality, since it is obviously false that these constitute our respective identities.

d) Opting out of Relation:

Failing to connect two utterances under the same topical content is an example of opting out of the relation maxim. If I suddenly utter that I will pick my interlocutor up at seven that night while being in the middle of narrating what happened during my holidays at the beach last month, I will be guilty of opting out of Relation, since there is no connexion between this utterance and my immediately preceding narrative.

These four examples all show that the result of the speaker's utterance communicates that at some point of the communicative exchange she has opted out of the operation of some maxim; their common feature is that opting out has been *overtly* communicated. When a speaker opts out, she does so by ostensively signalling to the hearer that she is unable/unwilling to comply with the requirements associated to the maxim at stake. What is interesting to note is that, as such content has been communicated, opting out is in this case the *result* of cooperative behaviour (in terms of CC and IC). Indeed, in order to infer that the speaker is opting out, both speaker and hearer will first have to cooperate in the exchange of this information. As a consequence, we can say that opting out is certainly CC cooperative in that the speaker is producing a content which, in order to go through, has to be interpreted as meaningful, but it is also usually IC cooperative, since the hearer is expected to interpret the specific meaning encoded in the utterance (which usually amounts to interpreting something like 'I have opted out/will be opting out of the operation of such and such conversational maxim').

Quite paradoxically, then, a speaker makes it clear that she opts out from the operation of the maxims by relying on these very maxims to get the message across; in other words, it is by

relying on the operation of the maxims that the speaker communicates that such and such maxims will not operate subsequently. This can come about either explicitly or implicitly, as is the case for instance when the hearer infers through the calculation of an implicature that the speaker is opting out. For example, if the speaker utters (15) when asked for specific information, she is actually communicating that she does not remember the particular information requested (cf. Martinich's example in (a) above):

(15) I am Jason Bourne[30]

Both examples are similar in that it takes some processing on behalf of the hearer to understand that the speaker is faced with a clash of maxims and that she does not wish to violate Quality (i.e. she does not want to provide false indications as to the information requested by her interlocutor), while choosing to opt out of Quantity. The difference with Martinich's example in (a) is that in (15) the interpretation is more complexly arrived at through implicature processing, that is, by a more elaborate operation of IC, which requires the exploitation of conversational maxims. What is interesting here is to see what happens on two levels of meaning: on the level of what is said, (15) contravenes to Quality (the speaker is not Jason Bourne) through the use of a metaphor, but this should be resolved quite easily by the hearer since it is manifestly the case that the speaker is not Jason Bourne. Now, on the level of what is implicated, (15) violates Quantity (since what the speaker is really communicating is that she is forgot the requested information) in order not to violate Quality (answering the question would amount to either providing false or not reliable enough information). Regardless of the intricacies of the example, what I wanted to stress here is that in the end, the speaker signals that she is opting out of Quantity, but that the way she intends to bring about this representation relies on the recognition, by the hearer, that she is making use of the maxims, in other words, that she is being IC cooperative. This shows that recognising that the speaker has opted out/intends to opt out itself may rely on IC both at the explicit and the implicit levels of meaning.

Opting out of the operation of a maxim can also meet the requirements of PC depending on the perlocutionary goal instantiated by the speaker. Think of a dispute where the proponent A wants her point adopted no matter what, and that, out of tiredness, her opponent B finally utters something like (16), thus making it clear that he is opting out of Quantity:

(16) A: Please tell me how it is that I was wrong in making that decision!

B: Ok, I don't know the whole story, so I concede that you were right.

[30] Jason Bourne is the main character of a cinematographic action trilogy about a special agent of a secret US military agency who suffers from amnesia and tries to recover his memory. Many thanks to Zoé Christoff for providing the example.

A, whose goal was to get B to admit that her decision was correct, manages to get B to PC-cooperate towards this goal, by getting him to concede his inability to continue the conversation. B's contribution can be taken to be a strict infringement of Quantity, since it communicates that B cannot provide a counter-argument due to his own ignorance of the facts, which is what A literally prompts B to do. Here, however, the infringement is actually consistent with the perlocutionary goal of the proponent; disposing of Quantity, in this respect, is instrumental to fulfilling B's goal and thus can be taken as a contribution A expected from B. Nonetheless, there seems to be a problem with this example: can we still call (16B) an infringement of Quantity when the conversational move supposed to be a infringement complies with the perlocutionary goal of the proponent? Or should we stick to the label 'infringement' in so far as, in abstract conditions, an infringement of Quantity is defined by a speaker's inability to provide sufficient information? Does this definition apply even when the conversation is not meant to continue? The problem we face here lies in the fact that the shared goal of maintaining the conversation about the topic of establishing whether A was right or not in making her decision is lost: somehow A replaced it with the goal of evicting the opponent from that particular exchange. In this last sense, (16) is fully PC cooperative because it meets the requirements imposed by A's perlocutionary goal of winning the argument.

(16) is therefore a clear example which shows that opting out of the operation of maxims can also result in PC, though this need not always be the case (when an interlocutor opts out without the perlocutionary goal of the speaker being fulfilled, for instance Martinich's example in (a), in which the speaker does not get the requested information). The difference between (15) and (16) could be explained in terms of a contextual difference: in (15) the exchange does not take place in a competitive mode while in (16) there clearly is one. Whether the participant's contribution will be PC-cooperative is thus a question which transcends the mere observation of maxims or whether participants choose to opt out from the operation of a maxim.

Beyond the specificities of conversational contexts, we could summarise our point on opting out of the operation of conversational maxims as follows:

- Opting out of the operation of maxims, as any ostensive conversational move, is CC cooperative to the extent that it supposes that the speaker make something mutually manifest. Opting out is successful only in the speaker manages to get the hearer to recognise her intention.
- The conversational move by which the speaker makes it manifest that she is opting out of the operation of conversational maxims, in itself, signals that a breach in cooperation will take place, since it is clear that the speaker is no longer able or willing to make a

contribution such as is required by the direction of the exchange she engaged herself in with her interlocutor.

- Opting out of a maxim therefore amounts, for a speaker, to saying "be aware that such and such maxim will no longer operate".
- The way this departure from conversational standards comes about is itself IC-cooperative,[31] suggesting that opting out from the operation of the conversational maxims itself requires an IC-cooperative contribution.
- Whether opting out is PC-cooperative does not appear to be determined by the contribution per se, but rather by the speaker's intention of abiding by the shared direction of the exchange.

It is striking that opting out of the operation of a certain maxim, as was the case in (16), may result from a clash between maxims, and opens Grice's categorisation to criticism, as the two categories seem to afford some overlapping. I now turn to assess a second way of opting out, namely opting out of the Conversational Principle.

I.2.1.2.2.2. Opting out of the CP

The second category of opting out phenomena Grice mentions lists cases where the speaker opts out of the CP. As hinted at earlier, it would be a mistake to consider that opting out from the operation of the maxims amounts to opting out from the CP, mainly because the operation of the maxims is concerned with meaning construction, while the CP is concerned with the conversational participants' rational behaviour in trying to maintain the conversation so as to stick to the mutually accepted direction of the exchange. The non-observance of a maxim is therefore not systematically IC-uncooperative since it may very well lead to a successful exchange of contents.

Now, opting out of the CP, strictly speaking, is denying the mere possibility of communication, which amounts to making it plain that CC will no longer hold. Grice (1989: 30) gives two examples in which it is clear that after the speaker's contribution, communication regarding the particular goal initially instantiated will cease:

(17) I cannot say more.

(18) My lips are sealed.

[31] This situation is apparently paradoxical, since it seems that non-cooperation requires cooperation, which reminds us of Chilton's (1987) 'CP paradox' (see §I.1.2.1.).

(17) and (18) communicate that the speaker is unwilling or unable to proceed with the conversational exchange, at least with respect to the conversational topic at stake until then, thus abandoning further CC *towards that particular topic* (it should also be said that since the exchange stops at that point IC – and most probably PC – regarding that particular direction of the conversational exchange will also be subsequently abandoned).

It is noteworthy that the two examples given by Grice, which signal that the speaker will no longer be CC cooperating, make use of the conversational maxims (since (18) violates Quality and (17), although in a perhaps too literal way were we to consider only the core meaning of 'can', also does). Thus, in order to announce that she will drop CC, the speaker is actually observing both CC and IC. Again, it appears that the state of affairs corresponding to the moment when a speaker has opted out from the CP is the *result* of an exchange where it is clear to both conversational participants that the speaker will no longer maintain communication on the topic at stake, if not communication altogether. That is why the characterisation of the phenomenon of opting out should receive special attention. A speaker opts out of the CP by relying on it, but there is a temporal interval between *communicating that one is opting out* and actually *opting out*, which indeed seem to be two different things.

When a conversational participant utters (17) or (18) she makes it clear that she is opting out of communication, or, more precisely, that she no longer wishes to utter anything else that would fit the accepted direction of the exchange. A particularly vivid illustration of a speaker opting out of the CP would be a child blocking his/her ears and singing very loudly so as to escape from the communicative exchange altogether.

We could consequently say that opting out of the CP amounts to abandoning the conversational exchange altogether (that is, abandoning the mutually accepted direction of the exchange, which can simply be translated into a refusal to proceed with the discussion of the topic at stake, for instance, or, at the other end of the spectrum of overt uncooperativeness, into a complete refusal to communicate any further). For Grice, cooperation is definitional of communication. As a consequence, a situation where conversational participants opt out of CC will not count as communication, even if at the time of the utterance in which the speaker announces that cooperation will not hold, both speaker and hearer are still engaged in a cooperative exchange (both IC- and CC-cooperative, the speaker formulating an utterance which complies with rational standards of communication, the hearer interpreting it in accordingly).

I.2.1.2.2.3. Opting out as a process vs. opting out as a state of affairs

If we want to properly assess whether opting out (either at the maxim or the CP level) counts as being uncooperative, we first need to break it down in two separate 'moments' in light of the

observations made above as to what seems to be the ambivalent nature of opting out, so as to get rid of imprecision in the characterisation of the phenomenon. On the one hand, opting out can be taken to be the meaningful communicative process by which a speaker makes it clear to all conversational participants that at a certain point she will no longer observe the standards of communication (either the operation of the maxims or the CP), or that she has not been observing these at some point of the exchange (e.g. in case she cancels an implicature which was drawn by the exploitation of a conversational maxim while she actually had opted out of its operation). On the other hand, opting out can as well be taken as the *result* of that particular conversational move, that is, the non-observance of standards of communication itself. In other words, we can either conceive of opting out as a process or as a state of affairs.

Opting out of the operation of maxims, when taken as a process (i.e. the conversational move announcing that the mutually accepted direction of the exchange has been or will be compromised by a refusal to abide by the requirements of the maxims) cannot itself be IC-uncooperative due to its communicative nature: in Grice's model, one must communicate that one will be opting out (or that one has effectively opted out) in order to successfully opt out. As the pursuit of a mutually accepted direction is the overarching goal of any conversation and since both speaker and hearer will be aware that the speaker is intending to opt out or has intended to do so, i.e., since both have understood that one of them was making it plain that s/he decided not to cooperate towards the particular direction of the exchange they were both engaged in until that point, any utterance announcing what will happen next or clarifying that the speaker has opted out at some point will count as IC-cooperative. In this sense, the conversational move of opting out is cooperative because i) it has been *communicated* (and therefore hopefully understood), and ii) both participants thus know what happened or what to expect.[32]

This suggests that in Grice's terminology, the fact that a speaker is opting out is necessarily *shared information*. It also suggests that a speaker cannot be IC-uncooperative in the process, that is, as she communicates that she is opting out / has opted out of the operation of the maxims.

[32] One may wonder how to describe, following this line of thought, a case where a speaker has opted out of the operation of maxims without having previously made it clear that she was going to opt out, which could potentially generate misunderstandings, as well as misleading scenarios. But this turns out to be a false problem, since when a speaker opts out of the operation of, say, Quality, while concealing this fact from the hearer, we are no longer in the category of opting out, but in that of unostentatiously violating a maxim (cf. §I.2.1.2.1.).

Now, the *state* in which a speaker finds herself and which has been announced by an utterance communicating that she is willing to opt out or that she in fact has opted out is itself *uncooperative*. If we think of it in terms of conversational stages, the state of having opted out (corresponding to the time interval in which maxim operation is actually dropped) is necessarily uncooperative, but this is a mere consequence of there being no communicative exchange whatsoever with regard to the previous direction that had been mutually accepted by both speaker and hearer. In other words, when one has effectively opted out, one is not cooperating in the exchange – that is, not cooperating towards the mutually accepted direction, which in many cases is to be understood as a conversation on one particular topic.

Opting out of the CP, in turn, equals opting out of communication *altogether*. It is plainly refusing to pursue the mutually accepted direction of the exchange. An utterance such as (19) would illustrate this:

(19) End of conversation!

Here the speaker makes her intention to opt out of the CP manifest quite explicitly, although this result can also be achieved by means of implicitly conveyed information, for instance by looking at one's interlocutor without saying anything.

Again, it seems that the conversational move by which a speaker makes it clear that she is opting out from the CP is cooperative: information is being transmitted, and this requires cooperation in Grice's model. But this is so only to the extent that we conceive opting out as a *process*. Now, the *result* of opting out necessarily turns out to be uncooperative, but this is by virtue of the fact that when a speaker makes it clear that she is opting out she *ipso facto* signals her refusal/inability to proceed in the mutually accepted direction of the exchange. That is, she communicates that she is no longer cooperating with regard to the said direction. There is in fact not even any question of being cooperative or uncooperative because once opting out is effective, communication relative to the initial mutually accepted direction is over. And as far as there is no communication, there cannot be cooperation in the CC sense. In cases where opting out concerns a previous utterance where the speaker had unostentatiously failed to observe the conversational maxims, whatever the hearer understood has to be put into question; for instance in cases of implicature cancellation, the speaker removes the communicative status of the implicature the hearer derived, with the result of excluding it from the meaning originally intended by the speaker, which could in turn be taken as a way of cancelling the communicative exchange regarding that particular content, so as to make as if it never took place.

To sum up, I have suggested that a distinction between opting out as a process and opting out as a state of affairs might be worthwhile our attention. The process of opting out is CC- and IC-

cooperative as any meaningful exchange of information while the state of someone having opted out is CC-uncooperative (and thus IC-uncooperative) because no communication (always regarding the relevant direction of the exchange) is taking place. Thus, only CC and IC are truly relevant in characterising opting out phenomena; as for PC, again, it seems that whether it obtains or not is irrelevant to the phenomenon of opting out, since one can opt out both in order to satisfy PC (when someone tells you to be quiet, for instance, thus getting you to opt out of CC), or in order to go against PC (as when I plainly refuse to answer a question, or reply with an overly evasive answer while letting it know to my interlocutor that I don't want to communicate at all).

It consequently appears that when a speaker opts out (either of the operation of the maxims or of the CP), the hearer will not be misled as long as this information is manifest. Moreover, we have seen that if we want to lean towards a conception of uncooperativeness as manipulative, deceptive or misleading communication, and at the same time try to find its links with maxim observance, it seems that the type of uncooperativeness opting out yields has little to do, if nothing at all, with a speaker being manipulative, simply because in Grice's spirit, opting out is *overt*, which makes it altogether cooperative.

I.2.1.2.2.4. Opting out, overtness and cooperation

Grice discusses opting out as an overt – i.e., endorsed by the speaker – way to fail to fulfil the maxims or the CP: in his words, a speaker may indeed "say, indicate, or allow it to become plain that he is unwilling to cooperate in the way the maxim requires" (Grice 1989: 30), which means that the speaker's uncooperativeness in opting out is part of the intended content of her utterance, and that she necessarily intends other conversational participants to be aware of it. In such a configuration, it proves difficult to link such uncooperativeness to our intuitive grasp of the particular type of uncooperativeness that seems to characterise misleading and manipulative communicative exchanges.

What this section on opting out nonetheless brings forward with regard to the broader topic of uncooperativeness is the importance of a criterion I will come back to, namely overtness. We have seen that according to Grice the category of opting out, either taken as a process or as a state, has nothing to do with a speaker concealing a goal she wants her hearers to unwittingly contribute to fulfil.

A pragmatic discussion on uncooperativeness must therefore be twofold: on the one hand, one can address uncooperativeness from the limited viewpoint of the strict non-observance of communicative standards, but on the other hand, such a discussion could very well start from a characterisation of the publicity of communicated contents, or, more specifically, the publicity

with which contents are intentionally communicated by speakers. In the first case, concluding that a speaker is being uncooperative is only a matter of checking whether she has observed communicative standards. This is certainly a technically clear-cut characterisation, but it relies on a very limited conception of non-cooperation, which straightforwardly abstracts from the conversational participants' mental states and focuses on the message itself and whether it complies with the maxims. Along these lines, we are left with a somewhat 'cooperative' conception of opting out, which is not of crucial concern regarding my topic of research, since there is *a priori* nothing misleading, even if impolite, rude, or otherwise annoying, in refusing to pursue a conversation.[33] Which leaves us with the second option, namely the idea that uncooperativeness as we envisage it has more to do with concealment.

Opting out, according to Grice, necessarily comes with a meaningful utterance expressing that the speaker will opt out or has opted out. A hearer will be aware that the speaker is opting out because he has managed one way or another to understand that it was her intention to do so. The cases of non-cooperation I am after are those where the hearer does not suspect that the speaker has intended to bring about in him such and such representation, or such and such action.

I am thus led to conclude that, in the case of opting out, failure to fulfil the maxims (and failure to observe the CP, CC- and IC-wise) cannot be taken as a characteristic feature of uncooperative/manipulative behaviour on behalf of the speaker, in that it is not misleading. In turn, this suggests that it is more relevant to a characterisation of uncooperativeness in communication to consider the criterion of covertness than the technical criterion of maxim non-observance.[34] In this respect, my position on failure of fulfilling the maxims is very much like Martinich's (1984: 36), who notes that "incidentally, one should not think that every violation of a Maxim of Quantity (or, indeed, of any maxim) is self-serving, malevolent, or surreptitious; the only requirement is that it be quiet and unobtrusive."

I.2.1.2.3. Clash between maxims

In Grice's original account, one particular type of failure to fulfil conversational maxims emerges from what he calls a clash between the operation of two or more maxims. Those are cases in

[33] Although we will see shortly that opting out can be part of a manipulative move.

[34] This obviously does not exclude the fact that maxim non-observance might in some cases prove manipulative or misleading (as in successful lies for instance, which violate Quality). But we are forced to admit that maxim non-observance is not a sufficient condition of manipulation; the reason for this has more to do with the fact that maxim non-observance is covert than with the non-observance itself.

which a speaker is unable to fulfil one maxim because it contravenes the operational requirements of another. A canonical example of a clash between maxims is the following (Grice 1989: 32):

(20) A: Where does C live?

B: Somewhere in the South of France

Here A is requesting precise information, which B fails to provide. Grice's explanation is that A assumes that B has no reason to be opting out and thus that B's less than informative answer is still cooperative. Assuming this, A is then led to infer that this must be because B does not know exactly where C lives and thus tries not to provide unreliable or false information. A is entitled to conclude that because B does not want to violate the maxim of Quality he uttered something which contravenes the requirements of Quantity (i.e. B did not provide sufficient information).[35]

According to Martinich (1984: 32-33), Grice was mistaken in considering this to be a genuine way of failing to fulfil conversational maxims on two counts. First of all, he argues that Grice mistook a *way* of failing to fulfil a maxim for a *reason* (or a cause) for failing to fulfil a maxim. It is because of an incompatibility between the operation of two or more maxims (in this case being precise would bring about untruthfulness, but on the other hand truthfulness would entail not being precise enough) that a clash emerges and consequently that a speaker is led to make a choice of preference between conflicting maxims. The second reason why Martinich thinks Grice was mistaken is that a clash of maxims is usually resolved by recourse to the ostentatious violation of one of the maxims. In this respect, this third category overlaps with the fourth one, namely *flouting* (see below §I.2.1.2.4).

With respect to the uncooperative status of instances of this third category – assuming in Grice's spirit that such a category indeed exists –, it appears that the result of a clash is CC-cooperative, in so far as B is genuinely intending to communicate something to A, and, what is more, that it is also IC-cooperative to the extent that the actual encoded meaning makes use of the conversational maxims (by an ostentatious violation of Quantity in example (20)). In addition, (20) is PC-cooperative, since by his utterance B signals to A that he is actually *trying to cooperate* towards one of the latter's extra-linguistic goals, namely to find out where C lives, even if the information B provides is not precise enough to fully meet A's expectations. Still, it is far more

[35] Notice the similarity between examples (15) and (20), which generate implicatures through a clash between maxims, which is an indication of the overlap, in Grice's account, between the categories of opting out and clash between maxims. More precisely, opting out could be interpreted as the result of a clash of maxims.

informative than a plain "I have no clue", or a "Somewhere on the European continent" would ever have been.

Having said this, there are no grounds to assume that a clash between maxims will automatically generate uncooperativeness. The only type of uncooperativeness we could encounter in such cases is very basic and limited to technical considerations. Under this configuration, uncooperativeness obtains only to the extent that a maxim is being ostentatiously violated; nonetheless, even the violation is intended to come across as the result of an intentional conversational move which relates to the mutually accepted direction of the exchange, which means that Grice's theory only takes into account *overt* and IC-cooperative maxim violations arising out of a clash. Coming back to example (20), if we focus our attention on the conversational move and its relevance to the mutually accepted direction of the exchange (i.e., finding out about C's place of residence), nothing in B's apparently uncooperative move is truly uncooperative at the end of the day, at least not from the point of view of a misleading/manipulative sense of cooperation.

I.2.1.2.4. Flouting

Grice defines flouting as a blatant failure to fulfil the conversational maxims. Flouting usually serves to communicate a content which has not been explicitly encoded, by giving rise to an implicature. Concretely, one may flout a maxim by deliberately and ostentatiously failing to observe its requirements. For instance, in example (21) below, B intends to implicitly communicate something like (21')

(21) A: Don't you think that that teacher is completely nuts?

B (seeing said teacher approaching behind A's back): So, where did you spend your last holidays?

(21') This is not a good time to talk about the teacher.

B's answer is completely irrelevant to the question formulated by A, and as such blatantly violates the maxim of relation at the level of what is said. However, according to the CP, A is *a priori* not justified in assuming that B is not cooperating in the exchange. A must therefore infer that B's contribution, because it is blatantly irrelevant to her question, implicates something else, namely in this case something like the implicature given in (21'). At the level of what is implicated, B's contribution is highly relevant not to A's contribution literally, but to the situation where it would be extremely relevant for A to know that the teacher she is insulting is around them at the time she utters the insult.

Here we encounter a feature I addressed in the preceding section (see §I.1.4.), namely the dynamic character of perlocutionary cooperation. In (21), it is clear that B's contribution is irrelevant to A's perlocutionary goal of complaining about the teacher on the level of what is said (considering that A wants B's opinion about the purported craziness of the teacher). But this does not mean that B did not understand nor that he is unaware of A's goal. In fact, it is because B considers it is an inappropriate time to adopt that particular direction in the exchange that he flouts Relation, at the same time as submitting an alternate direction for the current exchange, which is, on the level of what is implicated, very relevant to the situation at stake. In other words, in (21) B tries to take advantage of the essentially dynamic character of PC by at the same time implicating (21') and *ipso facto* proposing a new direction in the exchange. The success of that conversational move lies in A's ability to recognise the dual value of B's utterance. Example (21) therefore illustrates quite vividly the idea that every utterance is an occasion, for a speaker, to redefine the direction of the communicative exchange.

In terms of cooperation proper, it appears that (21) is CC and IC cooperative, since communication takes place and meaning is being transmitted according to rational standards.[36] In terms of PC, (21) illustrates two phenomena: that B is being PC uncooperative towards A's goal of complaining about the teacher, and the fact that B is asking A to cooperate towards a new – and irrelevant to the preceding goal – goal, which indicates that he is still seeking PC from A. In addition, B's implicature acts as the main motivation why A should PC-cooperate towards B's newly instantiated goal (a discussion about holiday venues).

The very same feature of the type of uncooperativeness Grice discusses, and which we evoked in preceding sections emerges here: maxim flouting is in this model always *overt*. As long as the speaker is making it manifest that she is failing to fulfil conversational maxims, she is fulfilling one of the overarching goals of any communicative exchange: that of being understood. This last observation results from definitional features of communication previously exposed, which require CC and IC. As far as PC is concerned, any analysis should state the perlocutionary goal at stake and which conversational moves relate to it. Only this way will we be able to precisely assess whether a contribution is overall cooperative or not. What can be said thus is that flouting a maxim *per se* cannot be a sufficient criterion to establish the uncooperative and misleading nature of a conversational contribution.

[36] We could argue that IC is operative at the level of what is implicated but not at the level of what is said; nonetheless, it is precisely because at the level of what is said (21) does not seem to abide by the CP that the hearer infers that (21) is cooperative only at the level of what is implicated. So overall, B's reply in (21) is cooperative, albeit with respect to the specific goal of avoiding a social gaffe.

A last word about the relevance of the category of flouting: example (21), in addition to illustrating the case of flouting maxims, can also be considered to fall under the category of opting out. As B fails to fulfil Relation, he also communicates that he disregards the goal of complaining about the teacher, which, in a way, amounts to his opting out of that particular goal. It could be reinterpreted as a simple "I do not want to proceed in the direction you requested me to pursue with you". Consequently, example (21) is also a case of opting out, even if the implicature one draws from (21') provides a justification of why B is opting out, which perhaps adds a layer of meaning which may make the conversational move acceptable. As we see, there is therefore quite some room to question Grice's categorisation, which is something Martinich had already noted, as did Thomas (Thomas 1995: 90).

I.2.2. INTERMEDIATE SUMMARY: OVERT UNCOOPERATIVENESS

So far I have tried to assess whether maxim violation is necessarily uncooperative or not, basing my discussion on Grice's seminal work. I have observed that his selection of apparently uncooperative cases of communication (leaving unostentatious maxim violation aside) is only concerned with *overt* phenomena, and that as such they are not part of the set of phenomena the present thesis is primarily concerned with. I drew this conclusion based on the assumption that any contravention of conversational maxims can be taken to result in uncooperative communication. It turns out, after examination, that overt maxim violation is not uncooperative in the sense I am interested in, and that there is consequently room to question the suitability of an attempt to use Gricean-inspired pragmatics – or at least there is room to introduce further refinements to the theory – in an account of manipulative and misleading communication.

As way of summary, this is what I have shown so far:

- The different ways of failing to fulfil the conversational maxims Grice identified (excluding unostentatious violations) result in non-cooperation only in a very restricted sense that can easily be disposed of; such non-cooperation denotes only technical violations of the maxims,[37] but cannot lead us to ascertain speaker uncooperativeness in terms of deception, for these overt violations are *manifestly* aimed at ensuring that meaningful communication still successfully takes place.

[37] One could inquire, then, about the proper uncooperative status of these violations. If the notion of uncooperativeness is to be taken in a technical sense, as I have exposed above, then any violation of the maxims is uncooperative. However, relaxing the constraints on the semantics of the term 'uncooperativeness', and considering that contravention to the maxims can still ensure meaningful exchanges, then the term uncooperativeness loses its technical specification and should specifically characterise intentionally and covert deceptive communication.

- More technically, we can say that maxim violation in Grice's theory is CC- and IC-cooperative by default (also in unostentatious violations, which can still count as successful communication)[38], because they partake in and make the exchange of meaning possible.
- The PC-cooperative status of utterances exploiting one of the ways of failing to fulfil the conversational maxims is perhaps more problematic, because it depends on other factors, such as the nature of the perlocutionary goal the utterance is intended to serve, the degree of manifestness of such goals as well as the assumptions conversational participants are able to make about their respective mental states.

Given Grice's original assumption that conversational participants are standardly cooperative in normal communicative exchanges, a discussion of what could be seen as uncooperative conversational moves, but which nonetheless ensure the success of meaningful exchanges, was in order. I have exposed the technical way in which Grice's work can be taken to consider non-cooperation in a strict sense (i.e. failure to observe the maxims and the CP), and more precisely I have argued that in this respect, non-cooperation is actually cooperative relatively to the goal of getting the meaning across, since maxim exploitation (including contraventions) contributes, for example in case of implicature derivation and specific figurative usages, to understanding full-fledged speaker meaning. The difference being that cooperation in those cases is operating at the level of what is implicated. It appears after this discussion that as long as these apparently uncooperative moves are *overt* they contribute to conveying speaker meaning, which in turn makes them neither intrinsically misleading nor manipulative.

Perhaps a fundamental question to be asked right now is the following: even while being overtly uncooperative (for instance by being sarcastic or by her use of metaphors), can a speaker still be manipulative? I'm thinking of cases where a speaker intends her hearer to derive an implicature which is misleading. *A priori* there would be no reason to rule out the existence of such instances; they could be put on a par with lies, with the difference that the latter operate at the level of what is said (covert violation of Quality), while the former would operate at the level of what is implicated. It would amount to letting the hearer infer, through implicature derivation, propositions which are truth-conditionally defective. This is a possibility I will discuss further along, but for the sake of clarity, let me first turn my attention to *covert* non-cooperation; I will return to this question after having discussed the latter.

[38] Covert violations still manage to get a meaning across, because the hearer interprets the utterance in a standard way; as a consequence, communication is still successful, and even in those cases the hearer uses conversational maxims to get to speaker meaning. This suggests that manipulation lies elsewhere.

The idea behind my exposition of Grice's four ways of failing to fulfil the maxims was to see whether they are connected with manipulation in a relation of determination. The question I set to answer was the following: are there grounds to consider that failures to fulfil the maxims necessarily result in manipulative communication? I have tried to systematically assess whether these failures inevitably lead to manipulative scenarios, and was forced to conclude that such is not the case. Leaving aside unostentatious violations, which require in my opinion separate treatment,[39] I can perfectly imagine that the remaining three ways of failing to fulfil the maxims can generate manipulative scenarios and non-manipulative scenarios alike, which is a clear indication that we cannot put forward any determining and stable cause-to-effect relationship from their occasional co-occurrences: it is therefore not because one fails to fulfil the maxims that one is manipulative.

Let us first take opting out. Imagine a scenario where my partner complains about me not putting the toilet seat down, and that I do not wish to discuss that particular issue, because I do not want to take the blame (or in fact for any other reason I do not want my partner to know about) but at the same time that I do not want my partner to think I am an evasive individual who refuses to take the blame. Imagine I reply by telling her that she is in a bad position to complain, given that she never takes the trouble to pick up her dirty socks, which she systematically leaves on the bedroom floor. If I manage to divert her attention to discussing the socks-issue, it can be said that I opted out of the CP in what regards the conversation about the toilet seat, since I did not take on what she intended us to talk about (i.e. her perlocutionary goal of blaming me). If I'm successful in launching my attack,[40] I will tentatively get away with my own concealed perlocutionary goal which consisted in avoiding the discussion about one of my faults. In such a case I have opted out, my opting out was moreover manifest, and at the same time it turned out to be instrumental to the success of my manipulative attempt, since I managed to distract my partner from her original goal without actually disclosing that such was my intention, notwithstanding the overt character of my opting out. I can thus conclude that one can in principle manipulatively opt out. Conversely, I have also shown above (cf. example (8)) that opting out can be completely PC-cooperative. The fact that opting out phenomena can lead to manipulative and non-manipulative scenarios indiscriminately demonstrates that, as far as opting out is concerned, we cannot take failures to fulfil the maxims as systematic indications of manipulative communication.

---

[39] See next section on IMT (§I.2.3.).

[40] This is a typical fallacy functioning as a red herring: the *ad hominem* fallacy, in its *tu quoque* variant. See section on fallacies for a development (§III.4.2.2.).

Cases of clash between maxims may also be found in manipulative and non-manipulative scenarios alike. Take example (20) again:

(20) A: Where does C live?

B: Somewhere in the South of France.

As noted above, this is an illustration of a clash between maxims, which is perfectly cooperative. But imagine now that B actually knows someone who knows the location of C's residence, but lets A infer that he doesn't know exactly where C lives (which is true), say, because B does not want A to pay C a visit. In that case, the example acquires a manipulative twist. It appears that beyond the speaker's failure to fulfil a conversational maxim, the manipulative tone of the example seems to result from an intention which B does not want A to know about. Once again, we are faced with an example in which the failure to fulfil the maxims (here, taking the form of a clash) is instrumental to manipulation, which is an additional indication that such failures to fulfil conversational maxims are not symptomatic of manipulative discourse.[41]

Finally, even if flouting is standardly used to generate implicatures in a covert way, there are cases where it can be used to deceptive ends. Imagine the following exchange:

(22) A: Can I trust Laszlo with my money?

B: Yes of course, about as much as you can trust Madoff.[42]

A would normally infer from B's response that B's ironic contribution is meant to communicate that Laszlo is not reliable enough to be entrusted with the management of financial assets, and therefore that A should not give Laszlo his money. This would be the standard perlocutionarily cooperative scenario: B is answering a request of A's about Laszlo's trustworthiness by being ironic. Now let's pretend that B has a concealed interest in A not letting Laszlo handling A's money, for instance because B wants to be asked to manage A's money himself. B's contribution in that scenario thus tries to disqualify Laszlo as a trustworthy person. The same utterance, taken in the same context, but this time supplemented with a hidden interest, becomes

[41] Of course, there is always room to argue that in addition to what I have noted about the example, there is an additional unostentatious violation of Quantity in B's contribution, since the information that B knows someone who knows where C lives would be relevant to A. By not revealing that particular piece of information, B is deliberately refraining from providing the amount of information requested. Although there indeed seems to be something of the sort happening here, I contend that the clash is still apparent, and that the hearer infers it as such., which in the end confirms the hypothesis that overt failures to fulfil the maxims can be instrumental to the success of manipulation.

[42] Bernard Madoff was the chairman of a financial corporation whose activities of wealth management originated what some people call the largest investment fraud in Wall Street history.

manipulative. This shows that the manipulative feature of the example does not depend on the violation of a maxim; rather, (22) becomes manipulative by virtue of B's intention not to disclose that he wants to be asked to manage A's money, regardless of whether Laszlo is a reliable money manager or not (Laszlo could be trustworthy or not without this affecting B's interest in A's money). With (22) we are consequently faced with two potential scenarios, both of which are concerned by a flout of Quality. The only difference between both lies in the fact that the manipulative scenario bears an additional concealed intention. From this simple example of flouting, we gather once more evidence supporting the claim that a failure to fulfil the maxims is not systematically synonymous with manipulation.

So far I have focused in detail on only three of the four cases where failure to fulfil the maxims obtains, leaving unostentatious violations aside. I have concluded that these may be regarded as cooperative as far as the overall conversational exchange is concerned because of their capital role in the exchange of meaning. Nonetheless, I have also shown that they may equally obtain in manipulative scenarios, which renders their usefulness as a criterion of manipulation worthless. In other words, what I have shown is that failures to fulfil the maxims of the three sorts detailed above are not necessary conditions of manipulative communication.

The fourth way of failing to fulfil the maxims – i.e. unostentatious violations – deserves separate treatment, which is something Information Manipulation Theory has chosen to develop, as we will see next.

I.2.3. DECEPTION AS UNOSTENTATIOUS VIOLATION OF CONVERSATIONAL MAXIMS

We have seen that when communication is successful, *overt* failures to fulfil the maxims comply with CC and IC; whether they result in PC is a far more complex question, which could very well turn out to be independent of the fact that they do not meet the requirements of the conversational maxims, because in principle PC depends on whether the utterance complies with the perlocutionary goal instantiated by the speaker and the participants' willingness to cooperate towards that goal, but not on conversational maxims observance (one can be PC-cooperative or PC-uncooperative independently of (non)observance of the maxims, as previously shown).[43] The important question to be addressed next is thus whether maxim violation has a significant role in *covert* breaches of PC, which, as I have gradually begun to expose, are the type of breaches that seem quasi definitional of deception and manipulation.

[43] This is actually the root of the problem any account of manipulation will face, which will be discussed further along (see §III.1.).

An original strand of research initiated in the 1990s frontally addresses phenomena of deception by taking on Grice's work and exploring one of the ways of failing to fulfil the maxims, namely unostentatious violations. Within literature stemming from the framework of Information Manipulation Theory (McCornack 1992a, McCornack *et al* (1992)) it is assumed that deception obtains when deliberate violations of conversational maxims occur in a *covert* way, that is, without the addressee being aware of them, as opposed to overt flouts, which are perceived as cooperative: "Verbal deception (by its very nature) is covert. Linguistic acts that *overtly* misrepresent information are perceived and processed as intentional 'flouts' of conversational principles (e.g., sarcasm), rather than deception (Grice, 1989)" (McCornack *et al* 1992: 25, footnote # 1).

This fairly recent empirically-grounded framework is meant to address precisely those cases of failure to fulfil the maxims which Grice dubbed "unostentatious violations", by considering the properties of verbal messages involved in communicative attempts to deceive.

I.2.3.1. Deception and Information Manipulation Theory (IMT)

Within theories of deception, it is usually considered that an individual attempting to deceive another individual must have a motive for doing so, and must aim at obtaining something from the latter (in the form of a belief or an action). Deception is thus an intentional endeavour:

> "There is probably only one characteristic of deceptive communication that is widely agreed on in the literature, namely, that an act of deception has to be intended by the deceiver. Messages that are unintentionally misleading are usually described as mistakes, gaffes, and the like." (Galasinski 2000: 18)

If we adopt such a conception, which in my opinion complies with common-sense, there is no way deception can be unintentional. It would be pointless to assume that a speaker can deceive a hearer without intending it.[44] If such were the case, then there would be no way of distinguishing, say, a lie, from a statement formulated in all good faith, but which encodes a belief which is false, the falsity of which is unknown to the speaker. A second aspect of deception is the fact that deceptive attempts must be covert, for if they were not, chances are high that they would not be successful (if I suspect that my interlocutor intends to deceive me, I would probably be very critical about whatever she is saying to me). The fact that deception is covert is

[44] My take, just as Galasinski's would be that unintentional deception does not exist and that it would actually denote cases of misunderstanding.

incidentally a clear indication of its intentional nature, for it would prove rather difficult to maintain that speakers can actually conceal things they are not even aware of.

These preliminary observations make room for a tentative account of deception couched in terms of Gricean pragmatics: first of all, Grice's model is designed to deal with intentional communication, which is the case of deception. Moreover, when dealing with failures to observe the maxims, Grice observed himself that some failures might be unostentatious, and that as a consequence the speaker, in those cases, might appear to be misleading (i.e., deceptive). This is precisely the aspect of Gricean theory that McCornack attempted to develop in two seminal papers published the same year (McCornack and McCornack *et al*. 1992), so as to ground a new theory of deception.

Information Manipulation Theory (henceforth IMT) thus tackles deceptive communication by strongly relying on the Gricean framework (Grice 1989) and assumes that intentionally covert contraventions (i.e., violations) of the conversational maxims are responsible for the generation of deceptive messages. IMT posits that it is by *covertly* violating Quality, Quantity, Relation and/or Manner that speakers manage to deceive their audience. As such, IMT is a theory which addresses the *features* of deceptive messages as covert deviations from conversational standards, and exploits what would be the 'uncooperative' counterpart of the Gricean model of communication as cooperation.

By considering that messages may be manipulated along four dimensions to meet deceptive ends, McCornack (1992) attempts to interface Gricean pragmatics and previous research on deception. Several deception researchers had already identified and focused independently on these dimensions: Ekman (1985), Metts (1989) and Turner *et al* (1975) attended to the amount of information given as well as its adequacy to truth (which, in Gricean terms, respectively corresponds to Quantity and Quality), Bavelas *et al* (1990) to the way messages are formulated (i.e., Manner), and Turner *et al*. (1975) to matters of relevance (i.e., Relation). McCornack takes that input and observes that it nicely accommodates with the Gricean maxims. Underlying his work is consequently the assumption that deceptive messages are "unique from other forms of discourse in that they involve deviations from what can be considered rational and cooperative conversational behaviour" (1992: 2). Unlike previous research which yielded problematic taxonomies of deception types,[45] IMT provides a consistent framework allowing to group deceptive messages under one (or more) violation(s) of the finite set of four conversational maxims Grice identified. IMT thus explores unostentatious violations of the conversational maxims, and assumes that these are responsible for deceptive communication.

[45] See McCornack (1992: 2-3) for the details of the critique.

IMT also relies on empirical testing in order to consolidate its main claims; as such, it is among the few approaches to deception that seek experimental validation of its main hypotheses. The study carried out by McCornack and colleagues first gathered a number of deceit-eliciting scenarios, which were then edited in order to avoid ambiguity and to get rid of potential biases.[46] A limited number of scenarios was then submitted to students, and these were instructed to supply messages in response to one of the potentially lie-eliciting situations; their task was basically to write what they would say in such situations. The results were analysed as actual deceptive data, and according to McCornack *et al*. (1992), empirical findings tended to confirm the main hypothesis according to which deception operates on the manipulation of information along the dimensions of Quantity, Quality, Manner and Relation.

From the pragmatic perspective adopted in the present work, IMT deserves thorough attention as an interesting piece of research on deception, mainly because it is an attempt both at bridging the gap between Gricean Pragmatics and psychologically-oriented research on deception and at grounding theoretical hypotheses on empirical testing. In addition, as Jacobs *et al*. (1996) observe, IMT is one of the first pieces of research which tries to make the connexions between linguistic pragmatics principles and the nature of deceptive message design.

As pointed out earlier, IMT frontally addresses *unostentatious violations* by strongly emphasising the importance of covert contraventions of the maxims. This allows the theory to test the features of the message itself; yet it also somehow constrains the scope of research onto a limited amount of possibilities, since there are only four maxims which can be violated. Furthermore, emphasis is put on the analysis of the production of deceptive messages and not so much on their reception; in terms of linguistic analysis, IMT therefore runs the risk of falling one step short of accounting for the whole communicative process. In the following section, I will expose the limitations of IMT and try to evaluate whether it can serve as a firm basis for a pragmatic analysis of deception, and ultimately non-cooperation (in the manipulative sense I am primarily interested in).

I.2.3.2. Limitations of IMT

Jacobs *et al*. (1996) criticize IMT after having replicated the experiment carried out by McCornack *et al*. (1992) and conclude the following: (i) IMT's main hypotheses are not

[46] Unlike in footnote #1, where I motivated my choice of the technical acceptation of the notion of bias as a tendency for mental processes to make errors and therefore to depart from otherwise habitually accurate and adequate information processing, here the term 'bias' refers to flaws in the experimental design that would compromise the validity of the data gathered.

confirmed by experimental testing as, according to them, data shows that deception is intrinsically concerned with violations of Quality, and (ii) IMT gives a very limited account of deception by failing to properly understand the Gricean framework. I will elaborate on both objections, which will lead me to discuss Jacobs *et al*.'s proposition as well, before adding further elements of reflexion destined to gauge the descriptive and explanatory value of IMT.

I.2.3.2.1. Problems with IMT: missing links between data and theory

Jacobs *et al* (1996) replicated the experiments carried out by McCornack *et al* (1992) and showed quite convincingly that the main assumptions of IMT cannot be confirmed by the selected empirical design. In this respect, it appears that the main problem with IMT stems from its rudimentary conception of deception which fails to capture all the complexities of the phenomenon.

Furthermore, and this is something Galasinski (2000) also notes, the experimental design chosen by McCornack *et al*. (1992) generates data on the basis of scenarios which elicit deceptive messages. In other words, we have no guarantee that the data corresponds to naturally-occurring or spontaneous deceptive utterances; what we get is a response to scenarios which potentially call for a deceptive message to be supplied. As Galasinski puts it, IMT tries to "analyse *elicited* 'deceptive' messages – that is, messages that participants *consider* to be deceptive, and this does not mean that they would actually use them in discourse in a natural setting" (2000: 33, author's italics).

While the idea that deception may take the form of covert violations of all four maxims is intuitively appealing, the problem with McCornack *et al*.'s account lies elsewhere, namely in the fact that an account based on maxim violation alone cannot capture the intricacies of deception.

Jacobs *et al*. (1996) postulate that deception is to be conceptualised as a phenomenon by which listeners are led to (i) form false beliefs, (ii) continue to hold false beliefs, (iii) drop true beliefs, or (iv) fail to form true beliefs. Accordingly, there is a crucial sense in which the outcome of deception will invariably have something to do with the parameter of truth, as deception is about ensuring that hearers either form false beliefs or dispose of true beliefs. This, in turn, suggests that deception involves some kind of manipulation of the maxim of Quality. Now, the subtlety Jacobs *et al*. introduce, and which IMT says nothing about, is that Quality need not be manipulated in the deceptive message itself as part of what is explicitly said: Quality may indeed also operate at the level of what is implicated. Accordingly, they suggest that deception always implies in this respect a violation of Quality, to the extent that deception generates the false

belief (including via a false implicature) that the speaker is being cooperative while she is not. This last point, though it seems crucial to an account of deception, is unaccounted for in IMT. [47]

While I agree with Jacobs *et al*.'s critique of IMT on the count of lacking an account of implicit deception, in all fairness, it seems to me that Jacobs *et al*.'s own position may nonetheless prove inadequate, notably concerning the idea that deception always involves a violation of the maxim of Quality for it prompts addressees to form false beliefs or to get rid of true beliefs. While I intuitively see the connexion between Quality and deception, I doubt whether this view is compatible with Grice's model. Manipulation of the maxims is something that happens at the production end of the communicative exchange since it bears on the way the speaker encodes the content into her utterance; the output representation the message yields is situated at the other end of the communicative process, namely reception. Consequently, the message itself, which can be 'maxim-manipulated' and the representation(s) such message yields in the cognitive environment of the addressee are two separate things. This is a distinction Jacobs *et al*. do not seem to consider themselves. They consider that deception always involves a violation of Quality since deceived addressees end up entertaining false beliefs or disposing of true ones. But we should be cautious in directly concluding that whenever an addressee ends up entertaining a false belief it is because there was a violation of Quality; some non-deceptive messages may yield a false representation, and these do not necessarily violate Quality (think of a situation where a speaker utters something true – maxim-violation-free – and the hearer infers something false; this is not a sufficient condition for deception to obtain, since the definition also covers misunderstandings). Maxim violation concerns the message the speaker produced, not its output representation(s). In other words, maxim violation is a property of utterances, not of mental representations, and just because the latter turn out to be false does not mean that there was a violation of Quality in the first place. It consequently makes no sense to characterise an output representation as violating the maxim of Quality (but Jacobs *et al.*'s account in fact affords such a conception). The violation is certainly responsible for the inadequate representation which is subsequently yielded; but the violation occurs somewhat *upstream* in the communicative process. So the claim according to which all deceptive utterances involve a violation of Quality is at best a mistake when we consider that the violation occurs on the output representation, that is, at the reception end of the communicative process. Jacobs *et al*.'s criticism of IMT does make a point on other issues, as we will see further along, but their account seems to display a misunderstanding of Grice's model as well, and should thus also be circumstantiated accordingly.

[47] See Jacobs *et al* (1996) for a detailed exposition.

Incidentally, the idea that all cases of deception involve violations of Quality can be disproved by examples of manipulation which exploit true information, such as the following (see Maillat & Oswald, forth.): on the 11th of September 2001, right after the World Trade Centre attacks in New-York, Jo Moore, British special adviser to the Transport, Local Government and Regions Secretary, circulated an internal email enjoining her colleagues to take advantage of the situation, as the events that had taken place earlier constituted a valuable opportunity to bury bad news.[48] As national and international public attention was quasi-exclusively focused on the terrorist attacks against the US, chances are that few people would have paid much attention to other news issued the same day, however disadvantageous to the British administration they may be. Moore's intention was to take advantage of the context of terrorism to disclose sensitive information, hoping that no one would pick it up. In such a scenario, the message provided would be true (for instance something like "The Secretary of State was condemned for tax evasion"), would not presumably yield false beliefs, and would *a priori* neither lead to the cancellation of true beliefs. In other words, this kind of manipulation does not exploit truth-conditional properties of the propositions conveyed in the message. Yet, it is still deceptive, insofar as it is part of a strategy meant to ensure that the disclosed information will not meet any opposition, which is something that would be likely to happen in a context where public attention is not monopolised by catastrophic world news. The example of the Jo Moore scandal demonstrates that deception does not necessarily, although it is most of the times the case, manipulate truth and falsity. As a consequence, violation of Quality cannot be said to be a necessary condition of deception either, which is something Jacobs *et al*.'s account would not be able to capture.

Coming back to IMT, Jacobs *et al*. (1996) contend that what seems to emerge from the data collected in McCornack *et al*'s (1992) studies is the idea that covert violations of Manner, Quantity and Relation are only deceptive to the extent that they result in addition in a violation of Quality, either at the level of what is said or at that of what is implicated. I tried to specify this claim earlier by considering that covert maxim violations generate representations which are problematic as they do not correspond to actual states of affairs, and not that they result in an overall violation of Quality, *contra* Jacobs *et al*. (1996). Now, if we conceive of deception as a phenomenon which fiddles with people's beliefs, then it necessarily has something to do with the adequacy of what is communicated to the actual worldly states of affairs. In this sense, violations of the maxims are instrumental to the generation of false beliefs. This means that a violation of, say, Quantity, is only deceptive to the extent that it prompts the hearer to infer

[48] See the *Guardian*'s coverage of the affair here (last accessed 15.08.2010): http://www.guardian.co.uk/media/2001/oct/09/terrorismandthemedia.Whitehall

something which is problematic with respect to the belief communicated, but the covert violation of Quantity *per se* cannot be taken to wholly capture what is going on as far as deception is concerned. Consider the following example, where a suspicious spouse asks their partner about their whereabouts:

(23) a. Did you go to that party while I was on my business trip?

b. Yes I did.

The speaker of (23b) answers the question truthfully, but does not disclose a crucial piece of information, namely that s/he did not go *alone* to that party. It would be highly relevant to the point of (23a) that the speaker of (23b) disclose that information, but instead, the latter chooses to violate Quantity by not revealing the entirety of what would be the relevant information. In this case and under standard conditions, the covert violation of Quantity triggers (if deception successfully goes through) the implicature that the speaker of (23b) went alone to the party (otherwise s/he would have said so). This precise implicature is what is problematic, because it leads to the formation of a false belief; in this respect, and as noted earlier, the covert violation of Quantity is only instrumental to securing deception by getting the inquiring spouse to adopt a false belief. But notice that (23b) does not contain any violation of Quality *per se*, as the statement is truthful.

What this example shows is that the whole point of deception is to secure that the audience will be forming a representation which does not correspond to an actual state of affairs (even if the state of affairs refers to something else than what is explicitly communicated by the utterance). Accordingly, an account of deception should go beyond the characteristics of deceptive messages in order to explain the effects maxim violations may trigger, and at the same time consider the other end of the communicative process, namely reception. I will elaborate on this last point in the next chapter.

Taking this into account, we see that one major weakness of IMT is that it does not take a required additional step to further the account of deception and that it does not extend the analysis beyond the stage of maxim violation. In refraining to do so, it fails to consider what the consequences of these violations are at the level of interpretation, which, in the end, is the level at which the outcome of deception will be decided (failure or success). In this sense, IMT falls short of providing a plausible account of deception because of its exclusive focus on message design.

I.2.3.2.2. IMT fails to understand the Gricean framework

IMT claims to make use of Gricean insights as it articulates its account of deception around covert violations of conversational maxims; however, and this might come across as a surprise given its reliance on Gricean theory, IMT does not make use of the central notion of implicature in its account. McCornack (1992) mentions that there is room to deal with implicatures within IMT, but considers that such issues are beyond its scope.

Yet, what is perhaps more troublesome is not so much the absence of an account building on implicature but rather the whole conception of communication underlying IMT. As Jacobs *et al* (1996) observe, IMT seems to reduce implicature generation to overt flouting and clash between maxims. This does not do justice to Grice's account, since in his model implicatures may also be generated by plain observance of the conversational maxims.[49]

In addition to this last comment, there is a further problem related to the conception of communication IMT seems to attribute to Grice. In Grice's model, the hearer's grasping of an implicature requires that he attribute to the speaker the intention of actually meaning the implicature to be processed. In other words, implicatures bear a communicative status in that their success (in terms of their being actually conveyed) supposes that the hearer recognises that they are intended. This is what Grice has in mind when he speaks for instance about blatant violations of maxims in his passage on flouting (Grice 1989: 30). Implicatures arising out of clashes between maxims similarly suppose that the clash is overt, that is, meant to be inferred as such by the hearer; furthermore, this is also the case for failures to fulfil the maxims or the CP which are typical of opting out scenarios. That means that whenever an implicature generated on the basis of a failure to observe a maxim is derived by a hearer, it is because the latter has recognised the speaker's *overt* intention to convey an implicit content. Accordingly, in Grice's theory, if a particular content is inferred but not attributable to the speaker as part of her informative intention, it cannot be called an implicature.[50]

Now, IMT, by positing that deception arises only from covert violations of maxims, closes the door to any consideration about potentially deceptive implicatures: there is absolutely no way, in Grice's model, for implicatures to arise from *covert* violations of the maxims, since, by

---

[49] In the Gricean framework, an utterance such as "Laszlo took a beer in the fridge and sat down on the couch", prompts us to infer via implicature the temporal succession of both eventualities by assuming that the speaker observed Manner, which prompts us to speak orderly. Thus, the observance of conversational maxims is a perfectly natural way to generate an implicature.

[50] A possible alternative some contemporary pragmaticians would resort to in order to denote those components of meaning which one cannot attribute with certainty to the speaker. (e.g. Blakemore 1992) is the notion of weak implicature.

definition, implicatures must be triggered by blatant – overt – failures to fulfil the maxims. In order to ascertain the legitimacy of an implicature, a hearer must be able, one way or another, to attribute to the speaker some sort of commitment to having intended the implicature to be derived. As a consequence, it is inconsistent to assume at the same time that deception arises from covert violations of conversational maxims and that implicatures may also be deceptive.

However, and this is where Jacobs *et al* (1996) expose a significant flaw in McCornack's theory, this does not mean that there is no way for an implicature to generate deception. Implicatures may communicate false representations, unbeknownst to the hearer, thus misleading him into entertaining false beliefs.[51]

Such a shortcoming is consubstantial to a rather simplistic conception of what happens in deceptive exchanges. As Jacobs *et al*. note, IMT assumes that "because deceptive violations of maxims are covert, listeners simply interpret what is said at face value." Instead, they argue that "what is deceptive cannot be at the level of what is implicated because nothing is implicated" (Jacobs *et al*. 1996: 72). And this is so because there is no way for a covert conversational move to be recognised as such and still at the same time achieve its intended effect; were it to be recognised, the intended effect would fail. To sum up, however unsophisticated IMT's take on deception may be, it will prevail as long as the implicit component of communication is not given proper attention, which is a shortcoming that IMT did not so far manage to settle.

I.2.3.2.3. Problems with identifying and ascertaining deception

I have noted above that IMT's focus on message design is limiting *per se*, in so far as the account establishes maxim violation as a definitional characteristic of deceptive utterances, somehow falling short of taking into account additional relevant factors. Even if most deceptive messages are probably conversationally faulty in a covert way, this is not obvious for every message displaying maxim violation. More importantly, were we to accept that maxim violation defines deception, what would we make of utterances where speakers merely make errors, that is, of utterances displaying *unintentional* maxim violations? For instance, if I utter:

(24) Geneva is South of Lyon

out of ignorance, since I happen to honestly believe that such is the case, does it make me liable to being deceptive? We can easily construe at least two different contexts, one illustrating a

[51] It is precisely because implicatures may be deceptive that a Gricean account – as any other pragmatic account of communication based on intention recognition for that matter – is challenged (see §III.1.2. below).

deceptive (24), the other a non-deceptive (24). Imagine first I am giving directions to Lyon to someone I don't like and whom I'd like to spend more time on the road than necessary. By violating Quality, I advise him to go in the opposite direction from where he is supposed to go, and thus manage to prejudice him by being deceptive. Now imagine that I utter (24) because I do firmly believe that Geneva is South of Lyon, while in fact such is not the case. The utterance is the same in both cases; the difference between them resides in the fact that in the first case it is accompanied by the intention of deceiving. Quite unsurprisingly, thus, we conclude that maxim violation alone is no clear indication that there is deception going on.

Of course, IMT would not support the claim that deception can be equated with sole contraventions of the conversational maxims. Rather, IMT posits that said contraventions are a necessary but not sufficient condition of deception. An additional necessary parameter of deception is covertness, which turns contravention of the maxim into a proper violation, i.e., makes maxim violation deceptive, since covertness requires intention.

Yet, the problem originating in IMT's exclusive focus on message design remains unsolved: while IMT can *describe* a difference between deceptive and non-deceptive utterances, it does not explain how it is that, though different, these may achieve the same result. It is one thing to describe deceptive utterances as covert violations of conversational maxims on the one hand, but quite another to explain how it is that these covert violations actually work; what we would need is a detailed account of why and how their covert nature determines their effectiveness. In other words, IMT provides a *descriptive* framework for the study of deception, couched in Gricean terms, but does not facilitate any research as far as the explanation of actual deceptive data is concerned, since the only prediction it will be able to form is that utterances displaying covert violations of maxims are deceptive. And this does not spell out i) the criteria allowing us to establish that the violation is *intentionally* covert, and ii) how and why the fact that they go unnoticed makes them successful.

Let us discuss more in detail the implications of an example such as (24). In case the hearer of (24) believes the speaker, whether (24) was deceptive or simply an error makes no difference at all as far as understanding – and to a certain extent being convinced – is concerned. Both deceptive and cooperative (though inaccurate) messages can yield the same effects, and the only difference between an 'ignorant' (24) and a deceptive one lies in the fact that in the second case it is accompanied by deceptive intent. But this may have no repercussion at all on the hearer's uptake of (24). Therefore, establishing that deception is a covert violation of a maxim does not provide any insights as to why (24) may be effective, based on observance or non-observance of maxims. Furthermore, the question we would like IMT to answer remains open: if, to a hearer, a lie or a false statement made in good faith trigger the same effect, how can we theoretically

explain the intuitive difference we do perceive between deceptive and non-deceptive messages? If not by looking at the properties of each separately, how can we draw a distinction between two phenomena whose effects can be the same?

What this question boils down to is basically to the uncovering of criteria meant to discriminate intentional from unintentional contraventions of conversational maxims, and moreover *covert* intentional violations from overt ones. But this IMT does not do. It is perhaps not a shortcoming of IMT's, but probably of Grice's model. Though his theory (cf. Grice 1989) accounts for different types of maxim violation with respect to the underdeterminacy of meaning and represents in that respect a revolutionary approach to the study of communication, it does not help ascertaining whether a content has been intended, which is incidentally probably not an exclusively linguistic problem. We sometimes have trouble identifying whether the speaker intended such and such implicature to be derived (see Thomas 1995: 88-89), which amounts to saying that we sometimes have trouble attributing intentions more generally. In the end, any judgment about an utterance's potential deceptiveness resorts to an assumption about the speaker's intention, more than to maxim violation, which seems to be an iffy criterion to support judgements about utterance deceptiveness. After all, there is always a possibility that the speaker made a mistake and took for granted certain information that was actually false. In other words, focusing on maxim violation to define deception proves insufficient because i) there is always the possibility that the speaker made an honest mistake, and ii) we have no sufficiently reliable solution to establish whether a particular violation of the maxims was intentional or not.

I.2.4. SUMMING UP: CC, IC, PC AND DECEPTION

To sum up, let me synthesise the different aspects of non-cooperation I have reviewed in what precedes in order to come up with concluding thoughts on the issue of the relationship between deception/manipulation and cooperation within a Gricean framework.

Communicative cooperation obtains in all cases of communication, deceptive or non-deceptive alike, as long as there is a successful exchange of information. Consequently, there is no point in considering that deception can be expressed in terms of non-cooperation at the CC level. Non-CC is not communication since it requires that individuals refrain from interacting with one another.

Informative cooperation obtains the moment a hearer has understood what the speaker meant, as interlocutors have made use of conversational standards of meaning management (i.e., conversational maxims and the CP) in order to end up entertaining similar representations. Non-

IC would thus depict a situation where the hearer has used these conversational standards to decode the contents of the speaker's utterance in a way which does not correspond to the speaker's own use of conversational standards in encoding her utterance. This definition affords at least two scenarios where IC does not obtain: misunderstandings – where the mismatch is accidental – and deceptive cases of communication, where the speaker for instance performs unostentatious violations of maxims.

In the first case, non-IC is independent of the participants' volition: at some point in the process of deriving speaker meaning the hearer has either mobilised inadequate assumptions or failed to use the conversational maxims the way the speaker intended him to. This is typical of cases where a hearer fails to capture sarcasm or irony, or more generally where he fails to derive an intended implicature. What is specific to these cases is that non-IC is not something deliberate; neither hearers nor speakers usually intend to misunderstand each other, and so non-IC in those cases is merely a *result* of a failed attempt to IC-cooperate. Hence, non-IC is not determined here by one of the participants' will not to IC-cooperate, because both do in fact do their best to IC-cooperate.[52]

Now, in cases of deception, it may be said that the speaker is deliberately not IC-cooperating. Maxim violation is by definition aimed at *covertly* failing to meet the requirements of the operations of a maxim. When a speaker lies, she violates Quality by uttering a content in support of which she lacks evidence or a content which she believes to be false, and if the hearer fails to notice the violation, he will be misled into entertaining a false belief. The speaker, by covertly refusing to go by conversational standards, is guilty of not-cooperating on the IC level, because she violates a maxim which is crucial to figure out that what she manifests as meaning, without the violation being ostensive, is actually not to be taken at face-value. More generally, any communicative situation where a conversational participant, speaker or hearer, is aware that a maxim has been broken but does not make it manifest, explicitly or implicitly, can be considered to fall within the scope of deceptive communication, whatever the motives of the participant are, and provided it would be relevant with respect to the overall exchange that all conversational participants be aware of the breach.

---

[52] Louis de Saussure (personal communication) points out that there are cases of deliberately concealed irony; I assume those to count roughly as weak implicatures, traditionally defined as implicatures that the hearer cannot take to be intended by the speaker with full certainty. In these cases, communication will presumably not fail, but it will not be 100% successful, as the hearer ends up with a less than certain representation of speaker meaning. But still, the hearer in these cases is at least capable of deriving two possible interpretations, a literal one and an ironic one, so one could still argue that those cases are IC-cooperative.

To the end of assessing whether deception can be equated with non-IC, I have taken Grice's failures to fulfil the maxims in detail and shown that opting out, flouting and clashes between maxims can translate into deception or non-deception, correlatively to the overtness of these failures. As it turns out, speakers may opt out, flout maxims and resolve clashes between them in a standard, non-deceptive way, as well as in a deceptive way. This shows that non-IC cannot be used as an exhaustive and thus reliable criterion to account for deception.

It consequently appears, as I have gradually tried to show through the preceding sections, that non-IC may translate into deception, although not necessarily (remember the example of the Jo Moore scandal). Considering aspects of IC will thus allow us to add detail to the analysis of some cases of deception but not to all, and as a consequence IC should be regarded only as a descriptive tool in an account of deception. It is one thing to describe what is happening when someone deceives other people, but quite another to explain how it works; in other words, qualifying deception as IC-uncooperative communication is not explaining i) the conditions under which non-IC occurs and ii) the reasons why non-IC might be successful. As far as the explanation of deception is concerned, my exposition of IMT's limitations accordingly suggests that an explanatorily adequate model of deceptive communication cannot be couched in Gricean terms, mostly in IC terms.

As a type of cooperation which encompasses far more than operational aspects of communication, perlocutionary cooperation probably deserves a different treatment, since it relates more to the goals one can pursue in and through communication than to technical aspects of meaning exchange. First of all, it should be noted that non-PC is not systematically deceptive: I can very well refuse to do my interlocutor a favour and explicitly say so without being liable to deceive. There is *a priori* nothing deceptive in refusing to be helpful, as long as the refusal is overt. It thus seems that we can speak of deception only when the speaker is *covertly* pursuing perlocutionary goals she does not want her interlocutor to know about.

Moreover, it very much seems that maintaining the appearance of PC is crucial to the success of deception. Probably one of the most problematic aspects of deception, were we to approach it through the lens of non-cooperation, lies in the multiplicity of deceptive goals one may come across: from an ethical point of view,[53] I suspect the motives of deceptive behaviour can range from good to bad. To take examples which can be found at both ends of the deceptive spectrum, I see no difference – and it is important that I highlight here that I'm only speaking from the

[53] Here I refer to the ethical dimension of perlocutionary cooperation Attardo (1997a) talks about as he formulates his Perlocutionary Cooperative Principle, which enjoins conversational participants to act as "good Samaritans" (see §I.1.4.).

perspective of the mere communicative process, i.e. considering somehow 'minimally' what speaker and hearer are doing in the communicative exchange– between a parent telling their child about Santa Claus in order to stimulate their imagination and a guru telling their followers that a supreme being will show up to announce a new world order sometime in the future, in order to collect personal benefit. Both scenarios involve a speaker formulating an utterance which they know does not correspond to a state of affairs in the world, thus misleading their addressees; furthermore, the success of both scenarios depends on the speaker's ability to conceal this fact from their addressees. Now, whether the deceptive messages lead to further ethically acceptable consequences in the real world is not the subject of this thesis. What is important here is that I assume the mechanisms by which addressees end up being misled in both scenarios to be roughly the same. It will be the purpose of a future section to develop an innovative account of these.[54] For the time being, I shall limit myself to noting that the nature of perlocutionary goals which speakers want to achieve through deception is not of utmost importance in an account of deception, perhaps because these relate to the speaker's preferences and desires, and above all because the contents of the latter cannot be captured scientifically. However, what can be said in order to specify a definition of deceptive communicative scenarios is that the speaker is pursuing a perlocutionary goal which *must remain concealed*, while manifesting apparent and acceptable PC. More precisely, I assume that the deceptive attempt would collapse the moment the hearer spots that the speaker is pursuing a perlocutionary goal which she does not want the hearer to know about.

Thus, a first tentative working definition of PC-uncooperativeness could be formulated as follows: an utterance is PC-uncooperative if the speaker, through the exchange, is pursuing a perlocutionary goal which she does not want her hearer to find out about. Hence, the hearer's ignorance of that perlocutionary goal is essential for the success of the deceptive attempt. Such a definition is advantageous first of all because it avoids unnecessary complications inherent to the notion of interest many researchers take to lie at the centre of deception (e.g. van Dijk 2006; Rigotti 2005; Saussure & Schulz 2005): saying that deception favours speaker interest necessitates a discussion about the types and natures of interests involved, and I have already hinted at the idea that this would take us on an unwanted track, as defining (dis)advantageous and (il)legitimate goals would require ideal omniscience and would additionally be compromised by ethical relativity.[55] The second advantage of such a definition is that it makes

[54] See chapter on contextual constraining (in particular §III.3. and §III.4.).

[55] See below (§I.3.1.2.) and Maillat & Oswald (2009) for an elaborated discussion.

room to address the means by which speakers manage to keep certain things concealed from their addressees in the communicative exchange, most notably their deceptive intent.[56]

The whole rationale behind my analysis of a tentative account of deception in a Gricean framework was to assess whether the exploitation of properties of cooperation for communication (IC) could enlighten us on PC-uncooperativeness or not. To deceive is covertly failing to PC-cooperate; the question was to see whether operational properties of communicative interactions would be exploitable in an account of deception. To this effect, I have successively considered whether the failures to fulfil conversational maxims developed within the Gricean model could provide helpful insights; it turned out that opting out scenarios, clashes and flouting situations ambivalently afford PC and non-PC, which is an indication that the Gricean framework is unsuited to fully capture a phenomenon as complex as deception.

I would like to bring up a final issue before concluding with this section. I stated earlier that it only makes sense to speak of PC relatively to a goal;[57] I have also noted that one of the problems any model of deception faces is finding a way to adequately account for the multiplicity of perlocutionary goals that can motivate deceptive communication. Given that cooperation requires two agents and a common goal, we could approach deception as uncooperativeness from two different angles: first, it could be said that the hearer, by being misled, unwittingly PC-cooperates towards the perlocutionary goal the speaker is pursuing. But is adopting a behaviour which happens to be consistent with a goal we do not know about really cooperation? I do not think so and as a consequence I do not think that one can say that misled hearers are PC-cooperating towards a goal they are not even aware of. Second, it could be considered that in deceptive scenarios the speaker is manifesting apparent PC towards a mutually manifest and shared goal. From that perspective, accounting for deception is about identifying the conditions which make the speaker's pretence to PC-cooperate effective.

Let me take an example in order to illustrate the two stances. Imagine Winston lies to his friend Laszlo by replying P to the latter's question (when non-P is in fact the case). Laszlo's perlocutionary goal is to get a reply to his question and if the lie is successful, he will believe Winston has PC-cooperated towards that goal, when in fact he has not. Deception, from this perspective, is indeed about manifesting apparent PC. From the perspective of Winston the deceiver, now, we could postulate that he has formed a perlocutionary goal such as "I want Laszlo to believe P", and that it is crucial for the success of that goal that Laszlo does not know

[56] This is also one of the directions of research Galasinski (2000) pursues.

[57] See §I.1.4.

Winston formed that perlocutionary goal. The problem, as stated above, is that we cannot say Laszlo is cooperating towards that goal, given that he knows nothing about it, even if his behaviour in principle conforms to such a goal. The crucial question here is the following: which is the relevant perlocutionary goal to characterise deception? If it is Laszlo's, the account should explain how apparent PC operates; if it is Winston's, then an account in terms of cooperation is compromised.

I have gradually reached the conclusion that the Gricean framework cannot fully answer the question of knowing how speakers manage to conceal their non-PC, because it only informs us on cases where IC is at stake while that deception may prove completely IC-cooperative, even if this is particular to specific instances of deception (e.g. the Jo Moore example). What this shows is that deception is fundamentally a matter of a speaker concealing one of her goals from the hearer, and that this concealment is necessary for the success of deception. In other words, the essence of deception, and manipulation, as Galasinski puts it, is

> "to affect the target in such a way that his behaviour/action is an instrument of attaining the goals of the manipulator, who acts without using force but in such a way that the target does not know the goal of the manipulator's action" (Galasinski 2000: 21).

In turn, this definition suggests that deception is foremost concerned with PC, which is common to all occurrences of deceptive communication.

The conclusions of the present section can thus be summarised as follows:

- *Overt* non-cooperation (CC-, IC- and PC-wise) is not deceptive.
- Deception can be defined as a *covert* failure to PC-cooperate. Whether it is IC-non-cooperative is a problem related to the means deception uses to be successful.
- IC and CC are properties of any communicative exchange in which conversational participants understand each other. Insofar as any communicative situation where meaning is exchanged operates on them, they cannot be used to distinguish standard from deceptive communication. Deception requires understanding, but not necessarily misunderstanding.
- While deception can be *described* as PC-uncooperative, the tools the Gricean framework has to offer cannot give the full measure of an *explanation* of deception, since there is no available criterion to directly assess the intentionality of deception.

In the next section, I will come back to the definitions of deception, misleading, persuasion and manipulation in order to delineate a working definition of each of these phenomena. This will

also lead me to extend the reflection beyond Gricean Pragmatics in order to assess whether pragmatic approaches to communication are suited to provide a scientific explanation of deception, so as to overcome the descriptive limitations I have so far exposed.

I.3. MANIPULATION

In this section I will first discuss some aspects of manipulation that we find across literature on the topic, in order to tentatively isolate the recurring features that meet consensus in different frameworks. I will also try to delineate the frontiers of the notion so as to distinguish it from neighbouring phenomena such as persuasion and argumentation.

I.3.1. DEFINITIONAL CRITERIA OF MANIPULATION: A SURVEY

Although in the previous sections I have already discussed the status of non-cooperation in pragmatic literature and at the same time identified some particular features of manipulative discourse (e.g., covertness and intentionality), I haven't yet formulated the working definition of manipulation I will adopt in the present work. Keeping in mind the underlying idea that manipulation is some kind of non-cooperation, this last section will thus be devoted to a discussion of the nature of manipulation, and its relationship with neighbouring phenomena such as deception, argumentation and persuasion.

Across disciplines, research on manipulation has usually been concerned with the nature of the manipulative message, its properties, and the social conditions in which manipulation is likely to occur. The purpose of what follows is to assess some of these accounts by trying to expose their limits, notably with respect to their characterisation of manipulative discourse. This will lead me to formulate a working definition, as I will explicitly try to identify some necessary and sufficient conditions under which manipulation occurs.

Addressing the 'what' question is consequently the main goal of this section, even if I will gradually shift my focus from the identification of manipulation to an account of the mechanisms by which manipulation comes to be successful, which is, in my opinion, a question that will prove far more interesting from a scientific point of view, as it allows envisaging the phenomenon in terms of relationships of cause and effect.

I.3.1.1. Truth

One of the ways of dealing with manipulative discourse is to use truth and felicity conditions as relevant criteria for a definition of manipulative strategies in terms of falsity and insincerity (cf. Rigotti 2005); the main assumption behind this view roughly amounts to considering that some features of the manipulative message, in terms of the content it encodes, do not comply with the

communicative standards of verbal interaction, notably with paradigmatic requirements of truthfulness and honesty.[58] This is the point made by Rigotti, who furthermore notes that in manipulation, "what is negative has to be somehow disguised as something positive", and that manipulation "twists the vision of the world […] in the mind of the addressee" (Rigotti 2005: 68). Rigotti's definition thus stipulates that the success of a manipulative move requires that the manipulative contribution seem ethically acceptable ("positive"): this is straightforwardly intuitive, as an individual would *a priori* not succumb to a manipulative attempt if s/he were to consider the propositions contrary to her/his ethics ("negative" being understood here as morally condemnable). Rigotti also suggests that there is something wrong with the conceptual content encoded in manipulative utterances, since manipulation prompts the addressee to form representations which do not correspond to actual states of affairs in the world. Typically, this makes room for lies to fall in this category, since it is crucial for a lie's success to be taken as a truthful statement. But note that such a conception also builds on the idea that there is something 'wrong' in the manipulative message itself, which is an idea we already found in IMT. In terms of the classification of phenomena, Rigotti's view entails that lying belongs to the broader category of manipulation and I would agree with him by contending that *lies are by definition manipulative*. I cannot proceed without devoting some attention to the notion of lie; let me thus first elaborate on the latter before I proceed to assessing whether criteria related to the truth of a message may help us in characterising manipulative discourse.

I.3.1.1.1. Lies

An intuitive definition of lying would state that lying is the production of false statements which are asserted as truths (i.e., a violation of the maxim of quality). But a definition centred on a violation of truth conditions should be carefully tested, since it cannot capture all occurrences which we would presumably call lies. Imagine I honestly, but erroneously, believe that Geneva is the capital of Switzerland, and that for some reason I want to make my Argentinean interlocutor, who knows next to nothing about Switzerland, to believe that Berne is the capital of Switzerland (which I believe is false), and that I utter (25) to this end:

(25) Berne is the capital of Switzerland.

The truth conditions of my utterance, *unbeknownst to me*, are actually satisfied in the world, because Berne *is* actually the capital of Switzerland. If we go by truth standards, (25) cannot

[58] In terms of Speech act theory, this amounts to saying that a manipulative utterance generally violates some felicity conditions of the speech act at stake, or, from a Gricean perspective, that it *covertly* violates the Cooperative Principle and possibly the conversational maxims.

therefore be called a lie. But I claim that this is wrong, and that (25) is in fact a lie. The reason for this is that, as far as *my* knowledge of the world is concerned, I sincerely believe that (25) is not true. In other words, I *believe* that (25) is truth-conditionally defective. Independently of my reasons for uttering (25), (25) is a lie, since it is produced with the *intention* of lying. Galasinski, elaborating on Bradac *et al*. (1986), explains this:

> "It is a particular set of a speaker's beliefs that constitutes a lie, rather than a relationship between the utterance and the extralinguistic state of affairs. The condition of the speaker's believing that her or his utterance is false is sufficient for the utterance to count as a lie. Also, the opposite is the case: a speaker cannot be held accountable for lying if the speaker says something that he or she believes is the case, although 'objectively', it is not." (Galasinski 2000: 23)

Although I find somewhat strange the idea of defining a lie as a set of beliefs,[59] the rest of the definition seems to make sense, since it captures examples such as (25). Further along, he adds:

> "Lying is related not so much to falsity of information but, rather, to the speaker's beliefs. It does not matter whether the statement is or is not false, as long as the speaker believes it to be false. It is the intention to lie and the speaker's beliefs about reality that are constitutive of lying. Lies are *statements* that the speaker *believes to be false* and that are *intended to mislead* the addressee." (Galasinski 2000: 97)

Reboul is on a similar note as she recalls Davidson to express pretty much the same thing:

> "It is generally thought that lying is saying something false. However, as was pointed out by Davidson, this is incorrect: 'It is sometimes said that telling a lie entails what is false; but this is wrong. Telling a lie requires not that what you say be false but that you think it false.' (1984: 258)." (Reboul 2002: 61)

Defining a lie as the act of asserting what one thinks is false also has the advantage of avoiding the inclusion of unintentional errors of production ('sincere' falsities which are due to either lack of or inaccurate knowledge of the world), the truth-conditions of which do not obtain, into the same category.

Once we acknowledge this first distinction, it appears that lying is about intentionally uttering a content which the speaker believes to be false. Whether the content obeys truth-conditions or not is irrelevant to the status of the lie (as seen with example (25)). This means that even in the case of lies, a criterion of truth compliance does not do the job, since its main requirements are i)

[59] I would probably define a lie praxeologically, as it is foremost an *act*, even if I do acknowledge the central role of the speaker's beliefs when she utters a lie.

the intention to deceive, and ii) that the speaker believes that the content she is communicating is false.

Making the definition of lies depend on intention and non-adequacy of beliefs to actual states of affairs opens the door to more complex issues occupying philosophers of language, semanticists and pragmaticians. Indeed, if we do away with truth-conditional meaning in an account of lying, then there is room to question the status of non-truth-conditional contents, typically implicatures. In order to illustrate how the discussion about implicit contents arises and how the question of their tentative status as lies can emerge, let us take the canonical example of scalar implicature:

(26) Ann has 4 children.

(26') Ann has *exactly* 4 children.

Is the speaker uttering (26) liable to have lied, in case she knows for a fact that Ann has 5 children? Truth-conditionally, there would be nothing wrong with (26), since if it is true that Ann has 5 children, by virtue of a logical implication, she *ipso facto* has 4 children. So by truth-conditional standards (26) is true. Now, if we do away with truth-conditions and go on to considering that lying requires *at the same time* the intention to deceive *and* getting the hearer to infer a false statement, then (26) is lying, for (26) is formulated with the goal of misleading the hearer into inferring (26') *and* (26') is additionally false (it is not true that Ann has exactly 4 children). Were we to classify (26) as a lie, we would thus need to reformulate the definition by stating that lying defines more loosely cases where the speaker *communicates* (either explicitly or implicitly) a representation she believes to be false.

This being said, I must admit I still resist the identification of (26), in the context where the speaker knows Ann has 5 children, as a lie. I obviously agree with the fact that (26) is misleading, but I cannot get rid of my intuition that (26) is not a lie, perhaps because as long as Ann has 4 children or more, (26) will be true. I would certainly concede that (26) is irrelevant in some way, that it definitely can be misleading, but not that it is a lie. If the speaker believes that Ann has 5 children, she must also accept as true the belief that Ann has 4 children by logical necessity. Let me put it somewhat differently: even if in (26) the belief which most accurately describes the actual state of affairs at stake (i.e. something like "Jane has exactly 5 children") is not part of the communicated content, what is communicated is still compatible with that belief. That is the reason why I would not consider (26) to be a lie.

It follows from this observation that the problem with (26) is not a truth-conditional one, but one of relevance: the speaker of (26) provides a representation which is not relevant to the goal of knowing exactly how many children Ann has, but which nonetheless is not false in any way. It

is irrelevant, and misleading in case it is taken to convey an accurate representation of the world. I would venture that the reason why this kind of example poses so many problems is precisely because we mistakenly confuse two dimensions on communication, namely truth and relevance. One can always be truthful and irrelevant at the same time, and in my opinion (26) is a vivid illustration of this, its misleading tone emerging precisely from the erroneous belief, on behalf of the speaker, that the speaker is being optimally relevant in selecting the utterance which best represents the state of the world.

A different way to conceptualise the notion of lie could be to restrict their scope onto explicit contents only by postulating that lies always involve material, or logical, contradictions between the propositions expressed literally ('what is said') and the epistemic status of these propositions in the speaker's mind. In other words, I will consider that (26) is not a lie because there is no contradiction between what the speaker knows is true ("Ann has 5 children") and what the speaker is explicitly saying ("Ann has 4 children"). There is a logical necessity between both propositions, since the former logically entails the latter. A contradiction certainly exists between the implicature (26') and the actual state of affairs (Ann has 5 children), but not between the explicatures of (26) and said state of affairs. It follows from this that a lie could be defined as the formulation of an utterance such that there exists a contradiction between its explicatures and a belief the speaker holds to be true. According to this definition, (26) is not a lie, because the explicatures of (26) are consistent with what the speaker believes to be true.

The next question to be asked naturally amounts to deciding whether implicit contents can enter the category of lies, which in my view can amount to the difficulty of deciding whether their defeasibility blocks this or not. This is typically a crucial question for any account concerned with manipulation, misleading and lying, since implicatures are particularly prone to yield confusion, misunderstandings and deception. So far I have no definite answer, because in the case of problematic implicatures such as (26') in the aforementioned context, what is at stake is no longer one single proposition, but two ((26) and (26')), and the glitch is that one yields a contradiction while the other does not. With (26), the speaker can always deny that she has lied because the explicit content she encoded is not contradictory with the actual state of affairs referred to. As I said above, I would say she is guilty of the charge of being irrelevant but not of that of being untruthful. Were we to call (26) a lie by virtue of the inadequacy of (26') with the speaker's beliefs, we would have to accept that (26) *exclusively* communicates (26'), and that the explicatures of (26) are completely irrelevant or that they should be disregarded from the

communicated content altogether.[60] This is obviously not a satisfactory stance; despite the fact that under normal circumstances the main point of these utterances is that the hearer infers the implicature, we cannot just flush the explicatures down the drain; they are still here, and they are not inconsistent with the world, as their truth-conditions obtain.

So the problem we face when we want to know whether certain implicatures can be called lies is a pragmatic one, as it exceeds the scope of truth-conditionality. Lies seem to require logical contradictions, and even if implicatures sometimes do yield such contradictions, they are not the only proposition 'packaged' within the utterance, as they can be accompanied by explicatures which do not yield logical contradiction (cf. example (26)). Utterances carrying implicatures are thus more complex to assess in terms of lying, because even if the relevant content intended to be communicated by the speaker (i.e. the implicature) does not correspond to actual states of affairs, chances are that their explicit content will pose no such problems.

So my take on the status of implicatures with respect to lying would be the following: implicatures are misleading at best, but a speaker communicating an implicature which encodes a truth-conditionally defective content cannot be held liable to lying as long as the utterance also contains explicatures which correspond to actual states of affairs. If this last condition is not met, then the utterance will generate false implicatures *and* false explicatures, this being a mere consequence of the inadequacy of the explicit content to the world. For instance, if I say "Ann had 6 children" when I know for a fact that she has 5, both the explicature and the implicature ("Ann has exactly 6 children") will be false. Yet, I contend that we can only term the utterance a lie by virtue of the fact that its explicit content and my belief about the actual state of affairs are inconsistent, and not because the implicature it conveys is false.

As a consequence, I do not believe that the category of lies by omission is relevant in an account of lying. Failing to provide information – thus misleading the hearer into forming representations which do not correspond to actual states of affairs – is not lying as long as *the explicit content does not violate truth-conditions*. To take a last example, if I am asked whether I cheated on my partner on the boat during last summer's cruise, and I answer no, I am not lying, even if I did cheat on my partner in a hotel on the coast while the boat was at dock.[61] I am certainly misleading my partner, and this is as much condemnable as a straightforward lie, but it

[60] *Explicatures*, in the RT tradition, are defined as explicitly communicated contents. Their derivation requires less contextual information than the calculation of implicatures; while Sperber and Wilson assume that all communicated information is inferential – and thus always requires the incorporation of contextual information – they contend that "the smaller the relative contribution of contextual features, the more explicit the explicatures will be, and inversely" (Sperber & Wilson 1995: 182).

[61] Many thanks to Zoé Christoff for the example.

is not lying, because it is in fact true that I did not cheat on my partner *on the boat*, which answers what was literally part of the interrogation.

I.3.1.1.2. Manipulation and lies

Returning to definitional issues, I am aware that considering the lie to be a particular case of manipulation may encounter resistance. Kissine for instance[62] considers that some lies may not be manipulative at all, since he considers that one can lie for the sake of lying, while one always manipulates to satisfy an ulterior motive (such as getting the hearer to unwittingly do something to the speaker's advantage). This position amounts to considering that lying can be an end in itself but that manipulation cannot; accordingly, manipulation is always a means to an end, whereas some lies may fall outside that scheme. If I agree with the fact that lying for the sake of lying is somewhat different from complex manipulations, in that in the former case the speaker is not necessarily pursuing any condemnable practical goal from which she could take advantage to the expense of the hearer, I see no solid reason to consider that the mechanisms governing the lie are different from those governing manipulation. Lying for the sake of lying is still an attempt at getting the hearer to entertain a belief which the speaker believes to be inaccurate. In this respect, getting the hearer to 'unduly' believe something, and getting him to do so in a covert way, is precisely the type of goal we find in every manipulative attempt. Whether the satisfaction of this goal is instrumental to the satisfaction of an ulterior or additional perlocutionary goal is in my opinion irrelevant to the characterisation of the lie as manipulative.

My reason for considering that lying and manipulating involve the same mechanisms also has to do with the scientific stance I speak from, and which I will elaborate on towards the end of this chapter. I will envisage manipulation from the perspective of the interpretive procedures it calls for, as I contend that the most suitable way to account for manipulation is to look at the type of processing it calls for, which I take to impose constraints misled hearers are not aware of. That is, I will look at things from the perspective of the hearer, and hence from the perspective of meaning construction. From this standpoint, I postulate that the way lies function is no different from the way manipulation proceeds: the hearer processes the stimulus by contextualising it in an appropriate but constrained set of contextual assumptions, the combination of which will yield an output representation through inferential mechanisms. Whether subsequent benefits are drawn by the speaker from that particular inferential process is a separate problem which only depends in my view on the particular perlocutionary goals the speaker entertains as she formulates her utterance. Thus, it is cognitive processing I will attend to, and in this respect, I

[62] Personal communication.

postulate that the way we process lies, when they are effective, is also symptomatic of manipulation.

Coming back to the issue of assessing whether truth-value may help us characterise manipulative discourse, I do not think that false statements are necessary nor sufficient conditions for manipulation to take place, which is something Rigotti (2005) also acknowledges. First of all, there is strong evidence in favour of manipulative mechanisms which do not involve lying or any falsity for that matter, as I have shown earlier with the example of the Jo Moore scandal (see §I.2.3.2.1). This is also true for examples such as (27),

(27) "Laszlo, there is no milk left for tomorrow's breakfast; could you go to the grocery store and buy some more?"

where the speaker provides a plausible justification for a request whose propositional content complies with an actual state of affairs; there is no milk left, and the need of milk for breakfast is a perfectly valid motive for formulating the request. Now, imagine that what the speaker really wants is to drive Laszlo away from home for a while in order to do something s/he does not want him to know about (e.g. preparing a birthday surprise, meeting a secret lover, etc.). In that case, (27) is unquestionably manipulative, but notice that this is completely independent from the truth or falsity of the information provided, since the truth of the both the formulation and the justification of (27) would be left unchanged should we consider a manipulative or a non-manipulative scenario. The difference between both thus seems to be a matter of intention, and not one of truth-conditionality, which is incidentally clear evidence in favour of the existence of manipulations which do not require falsity. At the same time, this heavily compromises the claim according to which manipulation has always something to do with truth and falsity.

Now, some scholars still point out that the manipulative status of a lie may be questioned: the case of parents telling their children about Santa Claus raises this particular issue, as highlighted by Saussure and Schulz when they argue that "it would be counter-intuitive that the parent telling the child about Santa Claus is performing manipulation" (Saussure & Schulz 2005: 2). If we are strictly looking at the nature of the message parents make manifest to their child by asserting Santa Claus' existence, truth-conditions are not met, among which existential presupposition. However, existential presupposition is also violated when parents tell tales about unicorns, and more generally, when a speaker's utterance is about fiction. Although it could be said that there is a difference between Santa Claus and unicorns in that children are aware that unicorns are to be found only in fairy tales, while they can reasonably be taken to believe in the actual existence of Santa Claus in the real world – to the extent that it might even have a tangible effect on their behaviour (being good all year in view of getting presents at Christmas) –, there is *a priori* no reason to assume that a criterion of truth is responsible for that

difference. Be it as it may, I would venture that any deliberate use of a false statement is manipulative as long as the addressee is not aware of it; that is, if we consider that the hearer's processing of the information turns out to be biased by virtue of the constraints the speaker imposes on the message.

As for the relationship between manipulation and lies, Parret (1978) argues that lying is limited to cases where a speaker communicates a false proposition which misleads the hearer into entertaining a new belief, while manipulation may function in several different ways: for example, the communication of a proposition may mislead the addressee into giving up a legitimate belief, into reinforcing a problematic belief or into preventing him from entertaining a legitimate belief, all of which can be achieved without resorting to lying on the speaker's part. These possibilities would, according to Parret, be difficult to capture either by the restrictions imposed by the very notion of lying or through an account solely based on the notion of truth. What this shows, beyond Parret's analysis, is that truth turns out to be an insufficient criterion when it comes to characterising manipulative communication, which translates into the idea that truth-conditional violation is neither a necessary nor a sufficient condition of manipulation.

I.3.1.2. Speaker interest

The notion of speaker interest has been highlighted by many researchers as a key feature of manipulation (e.g. van Dijk 2006, Rigotti 2005, Saussure & Schulz 2005). The claim is that manipulation is an intrinsically goal-oriented phenomenon designed to satisfy speaker interests. As van Dijk notes "manipulators make others believe or do things that are in the interest of the manipulator and against the best interests of the manipulated" (van Dijk 2006: 360). Saussure also highlights the connexion between speaker interest and the manipulative attempt: "communication is manipulative when the speaker retains some relevant information, or provides the correct information in order for the hearer to conclude that he should behave in a way which favours the speaker's interests, without being aware of it" (Saussure 2005: 119-120).

The notion of interest is intimately linked to that of goal. An interest can intuitively be defined as a goal that an individual might reasonably be taken to be willing to attain because it favours her/him in some way. Recall here that I postulated in the preceding sections on cooperation and non-cooperation that there are properly linguistic goals (i.e. somewhat 'local' goals pursued in the processing of information itself) and extra-linguistic – or perlocutionary – goals ('global' goals, such as the pursuit of happiness, well-being, receiving favours, etc.) in communication.[63]

[63] See Lumsden (2008) for a state-of-the art discussion of the distinction between linguistic and extra-linguistic goals from a neo-Gricean perspective.

Communication, including manipulative communication, cannot take place unless the message conveyed is understood by the addressee. In this sense, it could be regarded as one of the speaker's basic interests to intend that her message be adequately interpreted. But it would hardly make sense to use this acceptation of *interest* to describe manipulation, since the satisfaction of such interest is inherent to all communicative contributions. I therefore take it that when researchers discuss the issue of interest satisfaction in manipulation, they do so mostly with perlocutionary, extra-linguistic, interests in mind.

Now, several problems arise when examining the issue in further detail. First of all, if we want to define manipulation, as van Dijk does, as a phenomenon which runs contrary to the hearer's interests, we need to make sure that there are no cases of manipulation which favour the hearer's interests. In other words, the question we need to answer is the following: would it be possible for manipulation to favour the addressee's interest? Put more simply, is it possible to manipulate people to their own advantage?

Although this question might receive a negative answer at first glance, I argue that a positive answer should be preferred in a number of cases. Let me take an example: two friends, one of which suffering from a severe terminal illness, are having a conversation about medical treatments. Assume that the latter is phobic of needles and cannot stand the thought of being stung. Further assume that his friend just heard on the news that a new injection-based treatment which, in some cases, can cure the disease, and which considerably alleviates pain in all cases, has just been made available at the local hospital. What would we make of the communicative conditions if the healthy speaker manages to get his sick friend to go to the hospital without disclosing the nature of the treatment? It can hardly be claimed that this is an instance of hearer interest violation, to the extent that both participants in the exchange can reasonably be taken to pursue the satisfaction of the sick hearer's interest (i.e. to get better). Yet, I claim that this is an example of manipulation, notwithstanding the absence of hearer interest violation. And in this case, speaker and hearer interest do converge, at least with regard to the goal of improving the hearer's health. Of course, one could argue that in spite of this particular convergence, an asymmetrical configuration still remains here, in that the manipulative attempt runs contrary to the hearer's phobia of needles. Yet it seems to me that there is room to consider that the utterance was intended to fulfil a 'superior' interest; in other words, the success of this particular manipulative attempt is about subordinating the violation of the secondary 'avoid-phobia' interest to the satisfaction of the 'improve-health' interest. Here the manipulative goal is foremost to improve the person's health, not to trick him into getting stung in spite of his phobia, and so I would say that the real motivation of the manipulative attempt translates in this particular case into a convergence of speaker and hearer interest.

People do have conflicting interests, but this does not mean that manipulation is only about violating a hearer's interest: some violations, which in this sense can appear to be incidental, may happen in the pursuit of a goal which, in the end, satisfies one of the hearer's main interests. What I intended to stress with this example is that the consideration of *both* speaker's and hearer's antagonistic interests is not necessary to describe manipulation; examples such as that given above illustrate that manipulation, at least in principle, does not necessarily require the satisfaction of speaker interest *at the expense* of the hearer's interest. This counter-example casts doubt on van Dijk's take, according to which manipulation favours speaker interest *and* runs contrary to hearer interest. We could indeed always consider that manipulation runs contrary to at least some of the hearer's interests; however, contrary to a widespread take on manipulation, I do think that it is possible to manipulate people to their own advantage. This suggests that what makes a statement necessarily manipulative is not whether it is detrimental in any way to the hearer. Moreover, if we were to loosen van Dijk's criterion and assume that manipulation violates *some* hearer interest (whatever it may be), the feature would still lose its predictive power, as it would yield an overgenerating model (we can plausibly think that *some* interest will always be violated in human interaction).

A second problem with the operability of the notion of interest lies in its relativity. What may be considered as a legitimate satisfaction of interest for someone may not be considered as such by someone else. What is more, there seems to be room to question the legitimacy of the researcher's own judgment: on which grounds can we say that a researcher's judgment is accurate when it comes to assessing people's interests, which are for a large part private? These questions bring up the classical issue of the analyst's own (ideological) bias, and could definitely turn out to be problematic should the notion of interest become a central part of a systematic account of manipulation. What these two issues point to is that we cannot expect the notion of speaker interest to function as a definite criterion on a descriptive level in order to account for manipulation.

The idea is *not* that interest is completely irrelevant in a definition of manipulation, but that postulating that manipulation runs against the interests of the manipulated is not an accurate characterisation. Following Saussure's take according to which "to manipulate a human being may be about using a person, i.e. have that person adopt specific behaviours to fulfill the needs and interests of the manipulator, *regardless of the ones of the manipulated*" (Saussure 2005: 117, my italics), I thus propose that manipulation always favours one of the speaker's interests, *independently of the hearer's*; furthermore, the link between this interest, the utterance and the perlocutionary goal the speaker intends to satisfy must remain concealed from the hearer, since its discovery would jeopardise the success of the manipulative attempt altogether. The

advantage of framing things this way lies in the fact that it allows the inclusion into the category of manipulation of cases where speaker and hearer interests converge, which is something I claim might happen; to illustrative ends, let me recall here a scene from one of the *Batman* films where Bruce Wayne (the public identity of Batman) attends an official party organised in his honour; during the evening, he comes to know that the villain is present and that he intends to kill or harm the attendance of the mundane event. In order not to create a wave of panic by revealing to everyone that the villain is in the place and that something terrible is about to happen, he acts drunk and very loudly yells at everyone that they are vain and hypocritical, that their attending the party is only a matter of image, since he does not know them, and that they should all leave, as he declares the party over. As a result of the scandal he deliberately makes, people start leaving and look down on Bruce Wayne, who has actually just saved their lives. This example vividly illustrates in my opinion that people may be manipulated to their own advantage, and thus that the violation of hearer interest is definitely not a definitional feature of manipulative communication. If anything, it is the satisfaction of speaker interest which constitutes one of the necessary conditions of manipulation.

I.3.1.3. Covertness

A very intuitive and widespread view on manipulation – and which I have already postulated as a necessary condition of manipulation in the preceding sections – holds that in order to succeed, the manipulative attempt must remain covert. Quite straightforwardly, we could indeed assume that it is likely that manipulation will fail if spotted, as anyone noticing that their interlocutor is trying to get them to do or think something without being clear about their intentions will presumably adopt a very critical stance and increase vigilance. This is also incidentally one of the reasons why many authors qualify manipulation and deception as bearing some sort of conversational asymmetry (e.g. van Dijk 2006, Saussure 2005), which can for instance be expressed in terms of a difference in the amount of information possessed that is relevant to the conversational exchange altogether, and which is unequally distributed among conversational participants. Furthermore, there are reasons to doubt that manipulators would explicitly endorse their manipulative behaviour in communication. These preliminary remarks tend to confirm that it is necessary for manipulation to work to go unnoticed. This intuition is actually reflected in natural language, as noted by Blass (2005: 170), who claims that one can perfectly utter something like (28), but not something like (29):

(28) Can I persuade you to come with me to the cinema?

(29) ?Can I manipulate you to come with me to the cinema?

While it sounds perfectly natural to utter something like (28), the acceptability of an utterance such as (29), in which the speaker goes on record about her manipulative intention, turns out to be problematic, because covertness indeed seems to be crucial in manipulation (unless, of course, we take (29) to be meta-linguistically intended to produce some sort of humorous effect). The publicity of the manipulative intention consequently annihilates the success of the manipulative attempt.

Yet, if we are to assess covertness in manipulation, we should first be clear about its scope. A quick review of some elements that may remain covert in a manipulative message shows that a variety of things can be deemed to be covert:

- The overall manipulative intention: as noted above, the transparency of the manipulative attempt would seemingly compromise its success, making it crucial for the manipulative intention to remain concealed.
- Local and basic linguistic strategies: a linguistic device relying for instance on particular pronominal usages (e.g. us *vs*. them) could aim at surreptitiously establishing an in-group/out-group ideological discrimination, which in turn may influence further beliefs and behaviour.
- Global strategies: these strategies aim at creating adequate psychological and social conditions for the success of manipulation, such as controlled peer pressure or strategies designed to increase the faith in the speaker.
- Discursive strategies: a complex articulation of speech acts fulfilling a specific function, e.g. a fallacious argumentative move (see van Eemeren & Grootendorst 2004), may also require being covert in order to succeed.

We see that a great many aspects of verbal communication can remain covert. This complexity may compromise the operability of the criterion of covertness in an account of manipulation, insofar as different instances of manipulation may display different types of covertness. Such an enterprise runs the risk of being difficult to circumscribe in a model, given the heterogeneity of the set of variables. Some refer to psychological and cognitive phenomena (speaker intentionality), others to social issues (peer pressure, group behaviour), and others are about particular linguistic or discursive mechanisms (local and global strategies). In other words, the major difficulty we face in the applicability of covertness as a distinctive feature is a consequence of its problematic definitional range.

I advocate that the way out of the problem would be to consider that these multifarious aspects of covertness in communication can be addressed through an examination of the particular

processing the manipulative message calls for, since, in the end, their effect – or power – is only realised when an addressee processes the information conveyed by the manipulative argument.

Let me mention a further potential problem which emerges from psychological evidence (e.g. Barton & Sanford 1993). The results obtained through experimental designs meant to test the extent to which specific, crucial, lexical information is or is not accessed by subjects while interpreting a fairly simple narrative text revealed that even when they are warned about a potentially manipulative content (i.e. when they were told for instance that there are anomalies in the text), the anomalies would still go unnoticed, and the manipulative effect would still be achieved. This kind of observation would tentatively point towards the idea that covertness may not even be a necessary condition of manipulation, since in cases where you neutralise it by making addressees aware of the manipulative nature of the utterance, manipulation can still take place. However, the tension can be resolved by considering that in the experimental design the instruction did not explicitly alert the subjects about the *lexical* nature of the anomaly, and that this could explain that a significant proportion of tested subjects continued to miss the anomaly. It thus remains to be seen how a more precise warning would have led to the complete failure of the manipulative intent.

What seems to be certain, yet, is that manipulation indeed seems to rely on covertness. But the notion remains difficult to capture within a theoretical model and proves somewhat shaky to be operable in an explanatory account. What I am willing to grant for sure, though, is that the speaker's intention of manipulating the addressee, be it through a lie or through other means, is *always* covert, that is, in all instances of manipulation. This means that the covertness of the manipulative intention is a necessary condition of manipulation. However, cases where covertness is neutralised (such as the reported experiment of Barton & Sanford 1993) do not threaten the claim, as manipulation still takes place because addressees do not know what kind of anomaly to look for, and foremost because being warned about an anomaly is quite different from being warned about an overall manipulative intention. Arguably, if subjects were explicitly told to look for lexical anomalies, they would spot the incongruity. More clearly perhaps, if the respondents were told that the goal of the experiment was to see whether they would miss the particular anomaly or not, surely all of them would end up identifying the anomalous item, and the whole point of the manipulative attempt would probably go astray. I thus conclude that the covertness of the manipulative intention is a necessary condition of manipulation.

I.3.1.4. Social inequality

Research in Critical Discourse Analysis (e.g. van Dijk 2006) and in psychosocial science (e.g. Milgram 1974, Cialdini 2009) stresses the importance of the social conditions under which

influence is exerted and provides valuable insights into the underpinnings of manipulative communication. According to van Dijk, manipulators "need to satisfy personal and social criteria that enable them to influence others in the first place" (van Dijk 2006: 362), whereby it is assumed that asymmetrical social conditions of interaction are required for manipulation to take place. Typically, this is illustrated by an asymmetry between speaker and addressee(s) in hierarchical social positions (parent/child, professor/pupil, politician/member of the public, etc.), knowledge availability (highly educated/poorly educated people, scientist/non-scientist, expert/non-expert, etc.), access to public discourse, and so on. Underlying this view is the idea that manipulation exploits these forms of social asymmetry and their associated imbalance in the relations of power and domination to favour of the manipulator's goals.

While social parameters indeed contribute to successful manipulation, there are reasons to doubt that these are necessary conditions. They certainly make explicit certain conditions under which manipulation can be facilitated, but they are not powerful enough to capture all cases of manipulation. The justification for this is a shortage of compelling reasons straightforwardly excluding that manipulation could occur in the absence of prior social domination: in principle, I do not see why children could not manipulate their parents, why friends could not manipulate each other, why pupils could not manipulate their professors, etc. In addition, on the one hand a socially powerful individual or group may very well decide not to use manipulation when attempting to convince another subordinate group; on the other hand, manipulation can obtain within reverse social power relations. In other words, social asymmetry does not *necessarily* entail manipulation, as manipulation does not necessarily entail social assymetry. Van Dijk is aware of the difficulty but dismisses counterexamples on the grounds that they hinge upon personal psychological factors:

> "I limit my analysis to social criteria, and ignore the influence of psychological factors, such as character traits, intelligence, learning, etc. In other words, I am not interested here in what might be a 'manipulating personality', or in the specific personal way by which people manipulate others". (van Dijk 2006: 362)

It is interesting to highlight the fact that van Dijk both recognises and dismisses a non-social component in manipulative strategies since he identifies two aspects of manipulation, that is, its social and psychological underpinnings. However, it is clear that a suitable model for manipulative discourse should cater for both aspects.

The same quote also reveals that, even if he does not address personality criteria, his conception of manipulation is still *speaker-oriented*, for he associates the psychological feature of manipulation to the personality of the speaker. Even if it is desirable to include the psychological component of manipulation in a scientific account, the latter is not reducible to the speaker's

personality. I argue further that the crucial psychological aspect of manipulation lies in the way it constrains the addressee's interpretative processes, which is an issue concerning cognitive processing and which tentatively gives insights into universal mechanisms. In this respect, my proposal advocates that an appropriate pragmatic model of manipulation must shift its focus to and account for the cognitive (psychological) processes which underlie the interpretation of manipulative discourse. That is to say that my definition of manipulation will be *addressee-oriented*.

Incidentally, this does not rule out the relevance of social factors in the suggested account. Rather, such a stance will allow me to consider how these – together with other parameters – may effectively constrain the addressee's processing of the manipulative utterance. In other words, the social conditions of manipulation will not be approached as somewhat sociologically unified objects, but integrated as means to constrain information-processing.[64]

It therefore appears necessary to overcome van Dijk's restricted view of manipulation in order to be able to account for a wider range of cases. Indeed, contrary to his claim that "if manipulation is a form of domination or power abuse, […] it only makes sense to speak of manipulation, as defined, when speakers or writers are manipulating others in their role as a member of a dominant collectivity" (van Dijk 2006: 364), I argue that a more exhaustive account of manipulation is possible.

Manipulation is obviously not exclusively determined by social factors. Rather, an alternative take on the phenomenon could be that it is because people cognitively process information in a certain way that manipulation, on the social level, is made possible. It would probably be scientifically uninteresting to reduce an account of manipulation to mere social considerations of power and domination, and probably very limiting *per se*, as it would fail to adequately explain instances of manipulation where social relationships are irrelevant (e.g. cognitive illusions, which seem to work independently of social constraints, see e.g. Pohl (2004)). I will later on try to sketch out the lines along which a more comprehensive view of manipulative communication can extend van Dijk's account in order to host cases of manipulation that go beyond relations of social dominance, while at the same time capturing the influence that social relations do have on manipulative discourse.

[64] See also the section on social psychology (§II.2.).

I.3.1.5. Intention

It was highlighted earlier why, as most researchers (e.g. Galasinski 2000, Blass 2005, Saussure 2005, Rigotti 2005) contend, there are grounds to consider that manipulation is necessarily intentional. The gist of the argument was that since manipulation is necessarily covert, there is no point in assuming that a manipulator can conceal something s/he is not even aware of. And thus it makes no sense to speak of unintentional manipulation.

The main problem a scientific account of manipulation based on intention will face has to do with the private nature of intentions. I have already noted that the identification of a manipulative attempt requires that we be able to attribute the manipulative intention to the speaker. But this is far from straightforward, especially when we lack contextual knowledge that could corroborate the judgement of manipulation. Let me take an example.

Shortly after the 9/11 attacks on the United States and before the war in Afghanistan, George W. Bush said "You're either with us or against us in the fight against terror."[65] It would be tempting to consider that Bush's use of the fallacy known as the *False Dilemma* (presenting a choice between two options which are taken to be mutually exclusive and contradictory while in fact reality affords multiple choices) is manipulative in that it prompts hearers to infer – wrongly – that those are the only two available choices, thus leading them to take sides without a critical examination of the question – and such is indeed probably the case. However, we have no other evidence other than our analyst's intuition to back up our claim, because the use of a fallacy is not a systematic guarantee of manipulative intent. There is always the possibility that George W. Bush deeply and sincerely believes that the world is only made of good guys and bad guys. In other words, as long as we have no 100% failsafe way of ascertaining whether the speaker has such and such intention, we will not be able to provide conclusive evidence as to the manipulative nature of such and such utterance. Surely Bush is aware that the world is not organised in a Manichean dichotomy; but this is merely an assumption on my (and probably many other's) behalf. I have no way of ascertaining it and as a consequence my speculations about his mental states must remain charitable.

Thus, the identification of any manipulative attempt is always some sort of risky bet. Only in cases where we have additional information, for instance gained through historical research, multiplicity of sources, cross-referencing of discourses from the same person, etc. will we be able to gather evidence of manipulative intent. In this sense, it is quite reasonable to assume that Hitler manipulated the German people, but it is very risky to grant the rank of certainty to the

[65] A summary of the story can be accessed on http://archives.cnn.com/2001/US/11/06/gen.attack.on.terror/ (last accessed 04.09.2010).

idea that Nicolas Sarkozy, the French president, manipulates the French people in his speeches. We do have historical information in the former case that can corroborate our judgement, but we cannot claim the same for the latter – at least not now.

So the first problem with intention is the fact that it does not seem to be a reliable enough criterion to identify manipulation with certainty in a discourse. Yet, this is a significant shortcoming only if our aim is to *spot* manipulators, in other words, only if we are after means of identifying manipulation. But my primary aim is to understand how manipulation works, and from this perspective I assume that it does make sense to postulate intention as a central criterion.

An additional problem with the notion of intention is that, on the face of it, it could be greatly challenged when it comes to explaining cases of second-hand manipulation. These are cases where a successfully manipulated individual repeats the manipulative discourse without herself having the intention to manipulate. In such a scenario, her addressee might end up being manipulated, even if she is not trying to manipulate him. This example might be regarded as controversial in so far as its manipulative nature is concerned. Can we still consider it as a proper instance of manipulation? It could indeed be argued that we can, to the extent that the manipulated speaker, who propagates the manipulative speech, is relaying the original information, and thus it is as if the original ('first-hand') manipulative intention is indirectly relayed, though it is not intended as such by the relaying speaker. But would it make sense to speak of a manipulative attempt in this case? If the answer is yes, it would entail that there can be cases of manipulation with no manipulative intention, which seems incompatible with common intuition. On these grounds, one could assume that second-hand manipulation is not manipulation proper, even if the result of the conversation may tentatively be the same as the effect the original manipulative discourse had in the first place. I argue that there is no manipulation in this case, for the speaker is not concealing anything: the manipulated disciple of a cult sincerely believes what s/he is saying as s/he tries to make more people join the cause. But the fact that the result of the 'manipulative intent-free' discourse may be the same as the result of the intentionally manipulative discourse does suggest that the *effect* of manipulation does not directly depend on the manipulative intention, since it may be also achieved without it. But no surprises here: a belief P may be instilled in someone through different means (rational conviction, emotional persuasion, authority, coercion, seduction…). Yet, this is a clear indication that the explanation of *how* manipulation works must go far beyond the establishment of intention as a central criterion. Here we stumble again upon one of the major theoretical difficulties related to the study of manipulation: if we want to isolate criteria to *identify* the

phenomenon, then it seems that intention is central, but if we are looking for parameters which *explain* the phenomenon, then intention does is not crucial.[66]

A third issue which may turn out to be very thorny to settle relates to the difficulty in characterising the manipulative intention. Pragmatic approaches to communication usually describe communication as a matter of intention recovery: the speaker intends the hearer to derive a certain set of assumptions, and expects the hearer to infer that the speaker intends him to derive this set of assumptions. The problem we have with manipulation is that the manipulative intention does not seem to be of any of those sorts. First of all, it proves quite difficult to identify a purely manipulative content in the same way we can identify an informative content. There is no such thing as a properly manipulative meaning, since the same encoded content can be either manipulative or non-manipulative depending on the intention and the goals it is accompanied by. Second, the manipulative intention is meant to be concealed, and as such it bears no communicative status, since in order to be successful, the target must not recognise it. To take Starwson's terminology, who states that "an essential feature of the intentions which make up the illocutionary complex is their overtness. They have, one might say, essential avowability" (1964: 454), we must come to grips with the idea that the manipulative intention, due to its essential *non*avowability, is not part of the intentions making up the illocutionary complex, meaning that the manipulative intention cannot be envisaged as the type of intentions we find behind standard speech acts.[67] In other words, a manipulative utterance is not an illocutionary act because it is not meant to be recognised as such; Récanati notes in this respect that "a secret illocutionary intention is a contradiction in terms" (Récanati 1979: 98), thereby confirming the difficulty that I have expounded in the characterisation of the manipulative intention. These two observations suggest that the manipulative intention is very different in nature from the type of intention we can find in standard communicative exchanges; I take this to follow from the fact that the manipulative intention is not part of the complex of intentions playing a role in the processing of *ostensive* communication. This might constitute a serious handicap for an account of manipulation formulated in terms of a pragmatic theory.

For the time being, let me just sum up this first section on intention by insisting on the intentional character of manipulation. I have argued that we cannot manipulate people without intending it, since it supposes that we deliberately pursue a goal we deliberately do not want our

[66] A proper discussion of the theoretical options of a study of manipulation, as far as description and identification are concerned will take place in Chapter III (see §III.1.).

[67] See §III.1.1. below for a discussion of the problems raised by a Speech Act Theoretical account of manipulation.

addressees to know about. As a consequence, I will consider that all manipulative attempts are intentional, which makes intention a necessary condition of manipulation.

I.3.2. MANIPULATION AS NON-COOPERATION: TOWARDS A WORKING DEFINITION

We do find across literature a few definitions which seem to cover the grounds I have examined so far. For instance, both Galasinski (2000) and Blass (2005) go by Puzynina's (1992) definition and define it as "an attempt to affect the target in such a way that her or his behaviour/action is an instrument of attaining the goals of the manipulator, who acts without using force but in such a way that the target does not know the goal of the manipulator's action" (Puzynina (1992), quoted in Galasinski (2000: 21) and Blass (2005: 170)).

Saussure also points to all these aspects: when he notes that "manipulation of an individual is directly related to applying constraints, in particular constraints that the individual is not aware of" (Saussure 2005: 117), he stresses the covert nature of the manipulative attempt; he also acknowledges the role of intention, interest and even truth in some cases:

> "I notice that the notion of manipulation is not only associated with concepts like lie and intention but also the one of interest on the part of the hearer: a lie – or any defective statement – will be manipulative if the hearer utters it in order to have the hearer adopt a behaviour consistent with the manipulator's interest and possibly inconsistent with the hearer's own interest." (Saussure 2005: 119)

In order to cater for operational aspects of manipulation, Saussure also proposes that a manipulative discourse is one which is "produced in order to persuade the addressee of a set of propositions P1... Pn of type T with appropriate strategies" (Saussure 2005: 120). This calls for a specification of T and S, and much of his paper is devoted to this task. T is defined as truth-conditionally or truth-functionally defective utterances ("roughly, they are *wrong* in some way: false, unlikely, doubtful, inaccurate, inconsistent with the common ground; and therefore should be rejected by the hearer under normal circumstances" (Saussure 2005: 120)). As for S, the set of strategies, it denotes either local or global strategies, which encompass social and psychological conditions under which adherence to a message should be facilitated. While Saussure does acknowledge the role of covertness in the success of manipulation, the latter definition does not mention it; nonetheless, the definition is probably general enough to include the parameter somewhere within S, as the type of strategy employed to get the hearer to form certain beliefs or adopt certain courses of action must remain covert.

What is of great interest, first, is the mention of 'persuasion' in the definition. Some researchers indeed do consider that the frontier between both is very thin. Taillard (2000), for instance,

considers that persuasion can be covert, and that in those cases it becomes manipulation, as noted in Blass (2005). But probably Saussure's use of the term persuasion here is meant to convey the idea that manipulation prompts the hearer to depart from rational or critical reasoning as he processes the manipulative stimulus. I shall come back to the distinction between manipulation, persuasion and argumentation later on.

The second aspect of Saussure's definition which must be a focus of our attention is that it includes an operational dimension, and that as such it opens the door to a research programme centred on the different types of strategies a study of manipulation should account for. His definition not only provides criteria to identify manipulation, but also criteria allowing us to engage research on the explanatory level. In other words, Saussure's twofold formulation of the notion of manipulation as a phenomenon involving specific strategies allows us to envisage cause to effect relationships, which is the gist of scientific research. What this permits is to gain insight into the dynamics of manipulation from a pragmatic perspective. The remainder of the present thesis will largely be devoted to exploring this track.

I.3.2.1. A descriptive definition

I am now finally ready for the task of establishing the working definition of manipulation I will adopt throughout the present thesis.

Summing up what we find in available literature on manipulation, I have identified three necessary conditions of manipulation: intention, speaker interest (in the sense of goal pursuit) and covertness. Social inequality and falsity can characterise some manipulative attempts, but they are not definitional of manipulation, as counter-examples clearly show. I therefore propose the following definition:

(30) A working definition of manipulation:

An utterance is manipulative if it is intentionally used as a means to attain a perlocutionary goal the speaker is covertly pursuing.

(31) Corollary:

Covert perlocutionary non-cooperation is manipulation

(30) states that to manipulate is to produce an utterance which will – ideally – get the hearer to respond in such a way that the hidden perlocutionary goal of the speaker will be satisfied. (31) is meant to capture the notion of manipulation in terms of non-cooperation, specifically focusing on the perlocutionary level of conversation; a speaker manipulates a hearer when she gets him to unwittingly act in such a way that a perlocutionary goal of the speaker's, *which has not been*

mutually accepted, is satisfied. As I have shown in the previous section on non-cooperation, overt non-cooperation on the CC, IC and PC levels is not manipulative; what is manipulative is covert non-cooperation, IC and PC-wise.[68] This working definition subsequently includes IMT's conception of deception, but also makes room for further instances of manipulation which are not captured by IMT's (and by extension any Gricean) framework, such as cases which do not involve a violation of maxims at the level of what is said.

An additional difficulty arises from the term 'manipulation' itself, as it can apply to very different things in communicative scenarios: an individual, an intention, a goal or an utterance can all be said to be manipulative, which is perhaps why the literature on the subject seems to cover so many different aspects of the phenomenon, depending on the scientific discipline adopted to account for them. Yet, probably one of its central features is the fact that it is designed to satisfy a goal in a covert way, a goal that could not be attained without the manipulated person's naïve contribution. As a consequence, I will assume that the manipulative intention to covertly fulfil a goal is the definitional characteristic of manipulation.

At this point, we should make sure the proposed definition covers all kinds of manipulative communication. It seems to cover lies, as lies are indeed intentional attempts at attaining a covert perlocutionary goal (which can be as basic as getting the hearer to believe something which the speaker believes to be false). Misleading utterances such as (25) above, which are not lies by virtue of the inadequacy of their implicatures to the truth (and not of their explicatures) also fall under the scope of the definition I gave in (30), because they pretty much fulfil the same function as lies. The Jo Moore scandal, which I evoked in the section on IMT above (see §I.2.3.2.1.), also describes a situation where some people intend to covertly achieve a certain perlocutionary goal (having the public miss the disclosure of true information that is detrimental to the government). Presumably my definition captures all instances of manipulative communication, since they are presumably all designed to satisfy a goal of the speaker's which must remain covert and they prompt the hearer to behave in a certain way or to believe certain things which are instrumental to the satisfaction of said goal.

Perhaps one could oppose to this definition the case of second-hand manipulation, which I have already evoked above. To this objection, my reply is the following: I concede that the individual who unwittingly reproduces the manipulative creed does not intend to manipulate the hearer; this is unquestionable. Thus, according to my own definition, what some people refer to as 'second-hand manipulation' is not manipulation proper, which is something I can intuitively

[68] It makes no sense to speak of covert non-CC, as that would express the counterintuitive situation where a speaker would covertly manage not to display any disposition to communicate.

accept without too much resistance, as I have shown in what precedes. In so far as the relaying speaker does not intend to manipulate, even if she intends to convince, it is not manipulation. The original discourse was manipulative in the first place, i.e., it was designed and produced by a manipulative speaker who was covertly seeking to fulfil certain perlocutionary goals, but the relaying speaker reproduces it without the manipulative intention being there. In this sense, the relaying speaker, who once was the hearer of the original manipulative discourse and was indeed successfully manipulated, in principle is not aware of it as she reproduces the manipulative speech and this is sufficient not make her liable to manipulating.

Just because a manipulative discourse loses the manipulative intent when reproduced does not mean that it cannot generate similar effects. This idea has strong implications for the study of manipulation, as it points to another angle from which one can study manipulation, namely the way it functions. I have stated a couple of times earlier that my own account of manipulation will be concerned with the cognitive underpinnings of manipulation; it is precisely to deal with difficulties such as the ones raised by cases of so-called 'second-hand' manipulation that I will be led to formulate an alternative definition of manipulation meant to encompass most aspects of manipulation including the problematic ones just mentioned.

I.3.2.2. Explanatory issues: specifying the working definition of manipulation

As we approach a phenomenon from a scientific viewpoint, two issues are crucial and should imperatively be addressed: its description and its explanation. One must first provide criteria to identify the phenomenon under study. This is what I have tried to achieve through the formulation of (30) and (31) above. Then, the second component of scientific research must include the explanation of the phenomenon's behaviour in the real world: in other words the scientific endeavour must identify why and how it is that the phenomenon comes about, and this ideally takes the form of a predictive model designed to assess relationships from cause to effect so as to be able to understand it in its 'natural environment', that is, if we envisage the object of study from a naturalistic and mechanistic perspective. The variability and diversity of parameters relevant to human behaviour and interaction is in principle no obstacle for such a way of doing science: just because the phenomena involving human agency seem to be more complex than, say, falling apples, does not justify that the epistemology underlying their study must be different. The difference will be in the results the study will yield, as they will most probably reach the status of hypotheses instead of that of certainty. Yet, this is no problem if we go by Popper's falsificationist standards characterising scientific endeavours, who claimed that the purpose of scientific research was to formulate falsifiable hypotheses. Because in case they

are proven wrong, they still contribute to getting closer to the truth as they entail the elimination of one possible explanation. In his own words:

> "I can therefore gladly admit that falsificationists like myself much prefer an attempt to solve an interesting problem by a bold conjecture, even (and especially) if it soon turns out to be false, to any recital of a sequence of irrelevant truisms. We prefer this because we believe that in this way we learn from our mistakes; and in finding that our conjecture was false, we shall have learnt much about the truth, and shall have got nearer to the truth." (Popper 2002: 313)

Having opted for a hypothetico-deductive methodology in the vein of Popper's, I will also adopt a naturalistic and mechanistic approach to communication assuming that communicative phenomena can be studied as natural phenomena which correspond to states of affairs in the world; furthermore, such an approach tries to isolate the conditions governing their behaviour by considering, mechanistically, that these can be modelled into chains of cause to effect. The main purpose of the present thesis is the elaboration of such a model.

The case of so-called second-hand manipulation is of crucial theoretical importance for the explanatory dimension of an account of manipulation. It is interesting to note that the same discourse, whether accompanied by a manipulative intention or not, may have the same effects as both scenarios in principle can convince the audience. So, if the relaying speaker's discourse (which happens to be a reproduction of the original manipulative discourse) can be effective in spite of not being driven by a manipulative intention, it means that the message has specific properties that may be instrumental to the success of both the original manipulative effect and the sincere attempt to convince. The question we have to ask therefore is whether these properties are the same in the original and in the relayed message. I see no *a priori* reason to consider that such is not the case, especially if the original discourse is reproduced with a very high level of fidelity. It follows that the message itself must be given a particular place in the elaboration of the model, since there is probably something very particular about the it and the type of response it elicits from the hearer.

The crucial point here lies in a careful examination of what it means for a manipulative attempt to be effective, that is, to take Chomsky's expression, what it means for a manipulative attempt to 'manufacture consent'. The type of questioning such an inquiry calls for typically concerns explanatory aspects of scientific theories and goes beyond description. Hence an acceptable model of manipulation has to explain why the message, in the context in which it occurs, is manipulatively successful.

The definitions I gave in (30) and (31) above provide us with criteria designed to identify manipulation in terms of necessary conditions, all of which can reasonably be taken to constitute a sufficient one to identify manipulation. However, this does not quite tell us how manipulation works. Moreover, as mentioned above, the type of effect manipulation generates is not exclusive to manipulation. This highlights the need of a model that can adequately account for the processes involved in the success of the manipulative attempt; but since these effects may also obtain in standard conversations which are not manipulative while still meant to convince, also comes the need for an additional theoretical specification amenable to capturing the singularities of manipulative attempts. These can be assessed in turn by observing failed manipulative attempts and in particular failure to keep them covert.

One of the defining features of manipulation requires that the manipulative intention remain covert, so one of the tasks of the manipulator is to manage to conceal it. In order to understand how such a result can be achieved, we first have to understand the reasons that would make an addressee spot it. Quite intuitively, we could say that a manipulative intention is likely to be spotted if i) the addressee comes to suspect that the speaker is neither benevolent nor competent while at the same time trying to come across as such, ii) the message attempts to force the derivation of beliefs which are inconsistent with the addressee's own system of beliefs, iii) the addressee figures out more generally that the speaker expects something from him without disclosing her desire to have him behave in a certain way. Basically, an individual could be led to postulate the speaker's manipulative intention if he notices that the utterance he is processing is trying to get him to do or believe something while attempting to hide this fact. So a first component of an operational definition of manipulation, which echoes Saussure's (2005) typology of manipulative strategies, should define the manipulative attempt's ability to conceal its inconsistencies (be they logical or pragmatic). Consequently, I conclude that manipulation is in part about making sure that the hearer's attention is driven away from any type of inconsistency the message may yield, be it on the level of intention recognition or on that of beliefs. By *inconsistency* I broadly refer to incompatibilities between what the hearer may standardly expect from the speaker in terms of intention (benevolence, cooperation, competence, 'good Samaritanism', etc.) but also to incompatibilities between the content expressed both implicitly and explicitly on the one hand and actual states of affairs and the addressee's beliefs and values on the other.

In order to be successful, manipulation requires in addition that the hearer accept whatever content the speaker is uttering. By 'accept', I mean adheres to, is convinced by, finds relevant, obeys, takes up, etc., whatever the speaker manifests. This is an area of crucial interest for an account of manipulation, as it precisely concerns the cognitive mechanisms manipulators rely on

in order to satisfy their covert ambitions. Yet, it should be noted that this is a concern which is not exclusive to research on manipulation, because it is also the main object of philosophical and psychological research on belief fixation more generally, since it deals with the conditions under which individuals are led to accept beliefs as true. The example of so-called 'second hand manipulation' illustrates this pretty vividly, since the same discourse, independently of its being motivated by a manipulative intention or not, may yield the same results.

So the second crucial aspect of manipulative communication an operational model should cater for concerns the strategies employed to favour consent by spelling out the conditions under which manipulation comes to be successful. The type of answer I will provide to this question will take input, multidisciplinarily, from insights in cognitive science, argumentation theory, social psychology, linguistics and discourse analysis. My hypothesis is that what these fields of research have to offer can adequately be hosted in a cognitive pragmatic framework.

If we want to account for how manipulation works, we will thus need to account for processes of belief fixation and obtainment of consent, but also to specify how these come to be manipulative. Following what I've said so far, I will assume that the manipulative tone of consent obtainment lies in its covert nature, and that what makes manipulation different from other mechanisms of conviction is precisely the fact that it does not come across as such or that it does not disclose some of its goals which usually go beyond mere conviction. In sum, what precedes leads to an 'operational' definition of manipulation:

(32) Operational definition of manipulation:

> *A manipulative attempt (through communication) is successful when the addressee gives his consent to whatever is encoded in the content of the utterance while at the same time being prevented from spotting that such content is inconsistent with some of his assumptions (about the conversation or about the world more generally).*

The latter is the definition of manipulative communication I will use in the remainder of this thesis. Manipulative discourse thus takes the form of a twofold process: on the one hand it prompts the hearer to infer a specific content which he does not realise is problematic, while on the other hand it drives him away from spotting that that content (including the intention with which it was formulated) will be instrumental to the satisfaction of one of the speaker's covert perlocutionary goals.

Accordingly, the explanatory task now is also twofold: not only will I have to account for the conditions under which consent can be gained, but I will also need to account for the conditions determining the speaker's non-awareness of i) the inconsistencies the utterance yields, and ii)

the speaker's overall manipulative goal. An interesting parallel can be drawn here with Hamblin's discussion of the definition Aristotle gives of fallacies (Hamblin 1970: 12): while recalling that Aristotle defined fallacies as arguments which seem to be valid but which in fact are not, he notes that in principle two directions of research are accordingly possible: studying why fallacies *seem* valid and studying why fallacies *are not* valid. Manipulation makes its content *seem* acceptable, relevant, consistent, etc., while in fact proving *un*acceptable, *ir*relevant and/or *in*consistent. The definition given in (32) precisely aims at encompassing both directions, since it states that one should attend not only to mechanisms of consent obtainment but also to the reasons why manipulated addressees do not recognise that consent should not be given (which is crucial in determining why it is that the manipulative content seems 'acceptable'). I intend my account of manipulation to benefit from multidisciplinary insights precisely in that vein.

A last necessary comment concerns the difference made earlier between accounting for what manipulation is and accounting for how manipulation works. When I address the 'what' question by trying to characterise manipulation as a phenomenon that has occurrences in the real world, I study it as an object *from the perspective of the speaker*. In other words, I believe that characterising manipulation – i.e. defining its features so as to be able to identify it when faced with potential cases of manipulation – necessarily requires that we consider what it means, for a speaker, to produce a manipulative utterance. Unsurprisingly, thus, I have defined intention and covertness as aspects of manipulation of utmost importance. However, I feel that the account lacks informativity if we limit ourselves to looking at things from the perspective of the speaker, mainly because we can never be certain about speaker intentions, even if in communication we usually manage to recover quite accurately what speaker want us to recover (but which excludes manipulative attempts). As a consequence, I believe that gaining insights into the way manipulation works will allow us to apprehend the phenomenon more exhaustively, as we will be able to include in the account the cognitive aspects of information-processing which manipulation relies on in order to succeed. Hence, when it comes to answering the 'how' question, as we try to coin the conditions under which manipulation is successful, *the perspective must be shifted to that of the hearer*. Manipulative discourse is only effective if its addressees process it in such a way that it goes unnoticed; the answer to the question 'why can manipulation be effective' is therefore likely to be answered by looking at how information-processing is constrained. Ultimately, I believe that answering the 'how' question may even result in an enhanced definition and description of manipulation, for it forces us to consider that the success of manipulation heavily relies on the way the manipulative utterances are processed.

Having formulated these definitions of manipulation, we can now state the main focus of the present work. Since what seems a viable option is to consider that manipulation is a property of

utterances (even if by extension I will talk about manipulative intentions, behaviours and people), I will be primarily interested in the conditions which may make it successful, which amounts to taking a close look at the type of cognitive processing induced by these utterances.

In order to complete this theoretical incursion into the nature of manipulation, I will now turn to an exposition of neighbouring phenomena such as deception, persuasion and argumentation, before I conclude this first chapter by assessing the theoretical challenges a pragmatic study of manipulation faces.

I.3.3. NEIGHBOURING PHENOMENA

I.3.3.1. Persuasion vs. manipulation

A great deal of research in communication science, taken broadly, and social psychology concerns the notion of persuasion. As it intuitively relates to that of manipulation, it is of great interest to my own research to try to account for the difference between these two phenomena.

O'Keefe (2002) is probably among the most prominent scholars to address the subject. In his book, he starts by taking definitional issues into consideration. Although he does not believe that a scientific study of the notion of persuasion needs to coin a clear-cut 100% failsafe definition, he nonetheless tries to isolate common features in its paradigmatic cases:

> "Happily, it is possible to clarify a concept without having to be committed to a sharp-edged definition of the concept (and thus without having to settle such border disputes). Such clarification can be obtained by focusing on the shared features of paradigmatic cases of the concept." (O'Keefe 2002: 3)

The first relevant aspect that should appear in a definition of persuasion is the successful attempt to influence, whereby the notion of success is necessary to that of persuasion, as one cannot say something like "I persuaded him to come but I failed". In turn, success implies two additional features, namely i) the goal of persuading and ii) the intention to reach that goal. A fourth feature of persuasion is what O'Keefe calls "some measure of freedom (free will, free choice, voluntary action) on the persuadee's part" (O'Keefe 2002: 3), meaning by that that persuasion does not compromise the target's freedom to react to the persuasive utterance (one can willingly give in to or resist persuasion). In other words, O'Keefe considers that one is always aware of the persuasive attempt and consequently I take him to consider that persuasion is in principle always overt, or, perhaps more precisely, that persuasion does not necessarily require covertness. A fifth feature of persuasion is that it requires communication (and especially language): O'Keefe's gives the example of pushing someone from a roof top and claims that this cannot be an instance of persuasion, while talking her/him into jumping would be.

Finally, persuasion always involves a mental change in the hearer's mind (which O'Keefe notes is necessary to a change in behaviour).

Taking into account O'Keefe's overall characterisation of persuasion as "a successful intentional effort at influencing another's mental state through communication in a circumstance in which the persuadee has some measure of freedom" (O'Keefe 2002: 5), it appears that persuasion is very similar to manipulation, as the latter is also an intentional attempt at influencing an addressee. If we follow this account, the main differences between manipulation and persuasion thus appear to be that i) the latter is overt and consequently that ii) the addressee is free to resist giving his consent to persuasion due to his awareness of the attempt to influence, whereas these two conditions are not met in manipulation.

O'Keefe's definition of persuasion and the distinction I just made with manipulation seem to meet intuition; however, consensus is not achieved regarding this last point, since some researchers seem to consider that the terms 'manipulation' and 'persuasion' do overlap in a number of cases.

Galasinski considers that persuasion is "a subset of manipulative strategies, at the same level as deception" (Galasinski 2000: 21); moreover, he believes that persuasion is covert and plays on the addressee's adherence to values instead of justified beliefs:

> "Although a persuader sets out to induce a belief or attitude or, indeed, action on the part of the addressee, and does it also without the addressee's realising it, this is done by setting up a system of values to which persuadees are intended to subscribe. Targets of persuasion are made to believe that something is good or bad, rather than true or false." (Galasinski 2000: 21).

This vision cannot accommodate what Blass (2005) notes, namely that one can linguistically express one's endorsement of persuasion but not of manipulation. Also, Galasinski's take casts doubt on what I have identified previously as the overt nature of persuasion, or the fact that its overtness is no threat to its success. Furthermore, Galasinski seems to equate persuasion with the induction of values as opposed to that of justified true beliefs; this is symptomatic of the widespread view according to which persuasion plays on irrational and emotional aspects of reasoning involving values, contrary to argumentation, for instance, which is – ideally at least – supposed to involve rational and normative processes. A discussion about persuasion proper goes beyond the purpose of this thesis, but I would nonetheless argue, contrary to Galasinski, that persuasion may still concern the truth of propositions and that it should not be limited to

issues of value;[69] rather, what seems to me crucial to distinguish persuasion from manipulation is the fact that the first affords to be overt.

Taillard's (2000) account, anchored within a Relevance Theoretic framework and sketching out a model of persuasion combining insights from social psychology and cognitive pragmatics, also makes room for covert persuasion. She notes that the persuasive intention might indifferently be recognised or not without this compromising the persuasive effect altogether. That is, one can recognise the persuasive attempt and be persuaded, or one can fail to recognise it and still be persuaded. In the latter case, the hearer will not have expectations about what the speaker wants from him, since this is not part of what he inferred from the utterance, and thus whether he is persuaded or not will is up to him. In turn, this makes, according to Taillard, covert attempts to persuade liable to a higher risk of failure than overt intentional persuasive communication.

If we examine Taillard's claims in light of the descriptive definition of manipulation given in (30) above, it appears that what she calls covert persuasion I call manipulation; my take would be that the moment the speaker *covertly* intends the hearer to behave in a certain way (and in the scope of the term *behave* I include cognitive processes such as belief fixation) we are in a manipulative configuration. Let me take one of her examples in order to illustrate this point:

(33) "After reviewing research suggesting that customers tend to be more loyal (i.e. shop more frequently and spend more) to retailers who greet them personally, the management of a large chain of department stores decides to require that all sales associates address charge card customers by their last names. There are two possible strategies:

(i) Ostensive Communication: The policy is announced to customers as part of a new 'customer appreciation' package. A letter from the CEO is sent to all charge card customers stating that: 'At Smith's, we want you, our valued customers, to know how important you are to each and every one of us. To show this, we will always address you by name when you visit one of our stores.'

(ii) Covert Communication: No announcement is made to customers. In an internal memorandum to all sales associates, the CEO states the goal of the new policy: to make the customers feel valued without them recognizing that they are being treated in this way as a result of company policy. Associates should use

[69] In this quote from Dostoyevsky's *The brothers Karamazov*, "I cried out to you that it was he, he alone who had persuaded me that his brother Dmitri was a murderer!", it seems to me that the speaker was at some point indeed persuaded of the truth of the proposition "Dmitri is a murderer".

> customers' names, but under no circumstances should they comment on their reason for doing so." (Taillard 2000: 156)

Taillard argues that the second case is still persuasion and that its covert nature will deny its targets certain expectations that they would reasonably form in the first scenario. Now, I contend that if the strategy presented in (33ii) is later on revealed to its targets, chances are that these will feel deceived, whereas they probably would not feel the same way in scenario (33i). This, in my opinion, is likely to occur because the targets will have realised that a behaviour they adopted at some point was actually elicited from them without them being aware of the whole process. This is precisely where I would say the manipulative twist of the example lies: individuals are led to form a certain belief or engage in a certain course of action without realising that this is precisely what is expected from them, since in those scenarios speakers act covertly in the pursuit of their goals.

What emerges is that persuasion, at least from the viewpoint of the overtness of intentions, is standardly avowed; the moment it is covert it becomes manipulation. This is an also expressed, yet somewhat modalised, by Blass: "Persuasion according to Taillard (2000) can be overt and covert and it seems to me that the more it is covert the more persuasion becomes manipulative" (Blass 2005: 170). The idea of there being a continuum of overtness is questionable, to the extent that it proves difficult to conceptualise a gradable notion of covertness. Either one is aware or unaware of an intention; if I suspect there is a covert intention behind an utterance, it means that I cannot ascertain that it is there, and thus I cannot ascertain that it is overt. I think that the fact that sometimes addressees may feel unsure about whether such and such content had indeed been intended by the speaker is not to be equated with a continuum of this kind, but I will leave this discussion open for the time being. Thus, unlike Taillard (2000), I will consider here that persuasion cannot be covert, this constituting its main difference with manipulation.

I.3.3.2. Deception vs. manipulation

Galasinski's work (2000) provides valuable insights for the study of verbal deception, as it is to my knowledge the only piece of work explicitly devoted to the notion of deception in communication. He considers that "deception is a type of manipulation" (Galasinski 2000: 21), and that to deceive is to "induce in the addressee a particular belief, by manipulating the truth and falsity of information" (Galasinski 2000: 20). This is to say that deception is about getting the hearer to believe certain things, and that the means used to achieve this involve fiddling with the truth-conditions of the content expressed in the message. Galasinski notes that this last element is meant to distinguish deception from persuasion, as he takes persuasion to operate not on the level of truth and fact but on that of values.

According to Galasinski, the manipulative tone of deception is to be sought on the level of how it operates. While manipulation may operate independently of truth-conditional issues, deception will be likely to succeed only through the manipulation of truth-conditions. Deception thus necessarily involves nontruth, as it prompts its addressees to form a belief which results from the manipulation of the truth and falsity of information. In Galasinski's system, manipulation is envisaged as a discursive strategy, that is, as some kind of manoeuvre that the speaker deploys through communication in order to influence the addressee's beliefs and actions. In this sense, the properly manipulative aspect of such a communication lies in what the speaker *does* in the exchange, which brings out, though metaphorically, the original meaning of the term 'manipulation', i.e. the action of using, touching or operating an object. In this case what is manipulated are the properties of the message, in particular its truth-conditional features. It is thus no surprise that both persuasion and deception are specific types of manipulation in Galasinski's model. Manipulation, under this view, comes to define any discursive manoeuvre meant to covertly influence what addressees are able to infer from the speaker's utterance.

As persuasion and deception take the form of discursive strategies meant to influence addressees' behaviours and beliefs, they both have an incidence at the representational level; Galasinski proposes in this vein to construe manipulation as a higher-order concept of which deception and persuasion are particular realisations. More specifically, deception is about manipulating the truth-conditions of the message and persuasion about manipulating the values contained in the utterance, so as to prompt the addressee to form a problematic representation of whatever is being communicated (and of the world) in both cases.

Galasinski thus conceptualises deception as a particular type of manipulation which operates on truth-conditional aspects of utterances. Accordingly, manipulation is conceived of in his model as a type of constraint imposed onto the addressee's processing of information, which is consonant with the operational definition of manipulation I gave in (32) above. Manipulation can thus broadly denote any covert discursive manoeuvre meant to influence the hearer's beliefs or actions; whether it plays on truth conditions or on values will allow classifying it respectively as deception or persuasion, according to Galasinski.

There are no obvious reasons not to consider that deception is intimately linked with nontruth. I believe it is useful, in order to gain descriptive precision, to now refine the distinction between the two in light of Galasinski's observations, by characterising deception as a particular type of manipulative strategy which necessarily involves in one way or another nontruth (cf. Galasinski 2000: 21).

I.3.3.3. Argumentation vs. manipulation

Argumentation is traditionally conceived of as a rational – and what is more, ethical – means of convincing an addressee of the acceptability of a proposition. Van Eemeren & Grootendorst, the initiators and main proponents of one of the most influential contemporary model of argumentation, Pragma-Dialectics, define it as a "verbal, social and rational activity aimed at convincing a reasonable critic of the acceptability of a standpoint by putting forward a constellation of propositions justifying or refuting the proposition expressed in the standpoint" (van Eemeren & Grootendorst 2004: 1). By 'verbal', they highlight the communicative character of argumentation, by 'social' its interactive dimension, and by 'rational' its intellectual nature, which refers to the idea that argumentation involves inferential work on behalf of conversational participants, who are supposed to examine whether there are rational and reasonable grounds to accept the standpoint the arguer wants to convince her audience of. A philosophical criterion of critical reasonableness is used to determine argument validity (see in particular van Eemeren & Grootendorst 2004: 15-17), that it to say, whether the shift from the premises to the conclusion (which is the actual standpoint the speaker is putting forward) is acceptable.

If, just as manipulation, argumentation supposes that the speaker has the goal of getting the addressee to entertain a particular belief, which can in turn lead to an action, the way it achieves its results is yet very different for at least two main reasons. To begin with, the goal of argumentation is not covert, and neither are the propositions which constitute the argumentative complex, because it is part of the argumentative attempt that the hearer knows an argument is being presented before his critical judgement. It is in any argumentative move's interest that it be overt, because their crucial objective is to secure that the audience will accept the standpoint the arguments are meant to support. In other words, the success of the argument requires that the hearer reconstructs the argumentative path that leads from the premises to the conclusion; typically, this reconstruction is thought to play a major role in prompting for audience conviction. Only if the individual to be convinced is able to represent him/herself the inferential path leading from premises to conclusion will s/he be able to assess whether this path is valid or not. As a consequence, the overtness of the link between premises and conclusion, which is the gist of an argument, seems to be a necessary condition of argumentation. So a first difference between argumentation and manipulation, if we think of the descriptive characterisation I formulated in (30) above, lies in the fact that arguers are not pursuing any covert perlocutionary goal such that its covertness would secure its fulfilment.

A second point of divergence between argumentation and manipulation can be derived from the requirements of manipulation: I have hypothesised above in (32) that manipulation requires i) that the hearer adheres to the content of the utterance (in this it is no different from

argumentation), and ii) that the hearer is prevented from spotting logical and/or pragmatic inconsistencies. This last point is crucial when it comes to distinguishing argumentation from manipulation. As argumentation is about submitting a standpoint to critical evaluation, it ideally requires that every possible counter-argument or any potential objection be brought up, so as to establish the absolute validity of the standpoint. A sound and effective argument is meant to overcome every situation in which the standpoint could be jeopardised. Consequently, whereas the success of manipulation takes advantage of the hearer's being prevented from identifying 'disproving' contexts which threaten the manipulative attempt by revealing its inconsistencies, the success of argumentation requires that these contexts be accessed and ideally rationally overruled so as to establish the standpoint's tenability. Manipulation calls for the concealment of inconsistencies; argumentation fosters their representation, if there are any, in order to conclude in case it manages to overcome them that the argument can be tenable. An argument will be successful if it secures the tenability of the standpoint against a wide range – and ideally, the entire range – of tentative counter-arguments.

Having established this distinction, one could wonder if both phenomena may still co-occur. If we look at example (27) above, it appears that this is indeed an interesting possibility. This is because in (27), even if the argument of not having milk for breakfast expresses a perfectly sound motive for the speaker's request, the goal of the speaker's utterance is not directly to have Laszlo leave *because of the milk issue*: the speaker's real goal is to have Laszlo leave *tout court*, and it so happens that bringing in the milk argument *seems* a perfectly acceptable and moreover valid way of reaching that goal. What we can conclude from this example is that argumentation may indeed take part in a manipulative strategy designed to fulfil a manipulative intention. Were we to remove the manipulative intention from example (27), it would become straightforwardly argumentative.

It must be noted, furthermore, that if Laszlo ends up considering that the argument put forward by the speaker in order to sustain the request is relevant and acceptable, the manipulative attempt is more likely to succeed. This is one aspect of manipulation that is worth noticing: its success may very well involve the production and processing of what appear to be sound and acceptable arguments. It is precisely because manipulation involves an additional layer of intention that it can rely on lower levels of intention (such as the intentionality related to what is manifested by the speaker as an overt attempt to convince) in order to be successful.

I.3.4. CONCLUSION

Through this discussion of cooperation, non-cooperation and manipulation, I have characterised manipulative communication and reviewed the most salient definitional aspects that an account

of manipulation should cater for and have also exposed the main differences between manipulation and neighbouring phenomena. On the methodological side, I have concluded that the 'what' question was perhaps not as scientifically informative as the 'how' question and that an informed account of the latter might in fact provide significant advantages when it comes to explaining the underpinnings of manipulative communication. In this sense, I argue that the most important task in this endeavour consists in identifying the different types of constraints that make for the success of manipulation in communicative exchanges. It is the purpose of chapter II to discuss existing theories and accounts that have tackled more or less directly these issues so as to assess their contribution in terms of potential constraints underlying the success of manipulative communication.

CHAPTER II. A CRITICAL VIEW ON THE MAIN APPROACHES TO MANIPULATIVE COMMUNICATION, FALLACIES AND DECEPTION

"The principle that human nature, in its psychological aspects, is nothing more than a product of history and given social relations removes all barriers to coercion and manipulation by the powerful". Noam Chomsky

II.0. INTRODUCTION

This chapter will review literature in the Humanities and Social Sciences that relates to the topic of manipulation and deception. The scope of the review is fairly large as it will deal with discourse analytical research, argumentation theory, pragmatics, social and cognitive psychology; my choice of including all these is motivated by the conviction that multidisciplinary work is required to get an informed grasp of the mechanisms explaining manipulative communication, insofar as the study of communication itself requires insights from disciplines as varied as psychology, sociology, linguistics and philosophy.

The chapter is thematically divided. In the first section of the chapter, devoted to critical discourse analysis and argumentation theory, I take a look at how the issue of manipulation would be tackled from the perspective of discourse, that is, by taking the (verbal, textual) product of communication as the object of analysis. In doing so, I will be led to question whether it is possible to assess manipulative features of communication and to explain how it works through an analysis of its linguistic realisation. I will highlight what kind of contribution both paradigms can offer for the study of manipulative communication but will at the same time try to point to their limitations.

The second section of the chapter will be devoted to Social Psychology, in an attempt to shed light on the social constraints that can possibly play a role in the success of manipulative communication. I will refer to major works in the field and examine what kind of input, in terms of mechanisms of influence, can be gained for the type of pragmatic account I am after.

The last section of the chapter will review literature related to the cognitive underpinnings of information-processing, and in particular to cognitive theories which are designed to accommodate misunderstandings, errors and more globally all types of processing mistakes. The idea is to identify particular cognitive constraints on information-processing such as cognitive biases and heuristics that could enter the explanation of manipulative communication. The last part of the section will be devoted to presenting Relevance Theory (Sperber & Wilson 1995) in some detail, as it constitutes the cognitive pragmatic framework within which the model I will present in Chapter III will be hosted.

II.1. CONSTRAINTS ON DISCOURSE

II.1.1. CRITICAL DISCOURSE ANALYSIS

Among academic disciplines in the Humanities which focus on communication-related forms of domination between human beings is Critical Discourse Analysis (henceforth CDA).[70] One of CDA's overarching claims stipulates that language can be used as a means, among others, of producing and reproducing social inequality, and that the proper way to investigate this issue must address the intimate relationship between language, ideology and power. Accordingly, CDA practitioners aim to describe and explain these relationships, so as to raise critical awareness and ultimately help preventing this particular form of abuse, which constitutes in turn a guide for human action.

II.1.1.1. The critical stance: historical roots of CDA

The roots of this "research programme", as termed by Wodak (2006: 1),[71] draw on a heterogeneous body of research from the 20th century which has contributed to a critical scrutiny of how communication functions within society. Among them, it is worth mentioning French post-structuralist philosopher Michel Foucault, the movement of Critical Linguistics (initiated by Roger Fowler, Gunther Kress, Tony Trew and Bob Hodge), and the Frankfurt School (whose main proponents are Theodor Adorno, Max Horkheimer and Jürgen Habermas). In what follows I list some aspects of these works that contributed to the establishment of CDA as a research programme greatly devoted to the study of ideology in discourse, including mystifying aspects of ideology inculcation.

CDA practitioners quite generally acknowledge their affiliation to the works of Foucault, making use of his construal of power and his conception of discourse as a set of institutionalised meanings and values. It is assumed under this view that human relationships, in particular between social groups, are inherently marked by power and domination, whereby powerful people or groups of people are deemed to exercise some form of control over other people or

[70] For usual references in mainstream CDA, see Wodak & Meyer (2001), Weiss & Wodak (2003), Wodak (2006), Fairclough (1989, 1992, 1995), van Dijk (1998, 2008a, 2008b).

[71] Wodak uses the expression 'research programme' in a different sense than Lakatos' specific sense (Wodak, personal communication); in the latter, the technical term 'research programme' denotes a set of theories and experimental techniques surrounded by a protective belt of auxiliary hypotheses designed to prevent the falsification of an unchanging hard core of fundamental assumptions (i.e., a set of axioms, to take a mathematical analogy). Rather, Wodak uses the concept of 'research programme' in its non-specific acceptation, simply denoting a set of approaches sharing a common overall goal; the point in using the notion of 'research programme' is therefore contrastive, whereby it is assumed that CDA is not a homogenous theory.

groups of people. As van Dijk (1993: 249-250) notes, CDA addresses the issue of dominance, "defined here as the exercise of social power by elites, institutions or groups, that results in social inequality". Moreover, Foucault's conception of specific discourses as bound to specific socio-cultural contexts is compatible with the assumption that language is a vehicle for ideology;[72] this meets large consensus in CDA and serves as a founding principle of analysis. In a somewhat Whorfian way, it is indeed assumed that language operates specific categorisations of the world, and that these categorisations, which partake in the structure of beliefs, values and knowledge, are socially transmitted. Discourse, taken in this socio-cultural sense, constrains the actual discursive practices of individuals; a description of discourse, therefore, is taken to be a first step in the explanation of the propagation of ideology and the effects it yields on people's beliefs, values and knowledge.

The Frankfurt school of neo-Marxist critical theory is often mentioned as a major influence in CDA – Wodak & Meyer (2001) suggest that the use of the term 'critical' in the acronym 'CDA' can actually be attributed to the influence of this school of thought –, given that most CDA scholars adhere to the following idea, taken from Habermas: "language is also a medium of domination and social force. It serves to legitimize relations of organized power. Insofar as the legitimizations of power relations, …, are not articulated, …, language is also ideological." (Habermas 1967: 259, quoted in a number of CDA publications, e.g. Wodak (2006: 5) and Weiss & Wodak (2003: 15))

Drawing on the Habermasian idea that societies should foster 'ideal speech situations' in order to achieve collective consensus so as to bring about appropriate courses of action – and that collective consensus can strategically be exploited for domination purposes –, CDA tries to deconstruct the linguistic devices and strategies that powerful social actors use in order to control other groups. Following the idea that language is a means of bringing about the satisfaction of a specific social group's interests through the dissemination of ideology, it is believed that the deconstruction of the way ideology is communicated by specific linguistic strategies will enable individuals to resist, and ideally fight, social domination. Such is the avowed goal of CDA, which is designed to "investigate critically social inequality as it is expressed, constituted, legitimized, and so on, by language use (or in discourse)" (Weiss & Wodak 2003: 15) so as to "demystify discourses by deciphering ideologies" (Wodak 2006: 4).

According to Chilton (2007), the term 'critical' was first explicitly introduced in the academic field of discourse studies within Critical Linguistics (CL), as a way of characterising a particular linguistic stance. CL is a movement which emerged at the University of East Anglia in the late

[72] See O'Halloran (2003: 11-12) for a discussion of the Foucauldian influence in CDA.

1970s under the impulsion of Roger Fowler, Gunther Kress, Tony Trew and Bob Hodge and which has undoubtedly set the path for the type of linguistic analysis performed within CDA. Their seminal book *Language and Control* (1979) was among the first to focus specifically on the ideological aspects of language, which George Orwell had pointed at in his novel *Nineteen Eighty-Four* (1949). Accordingly, critical linguists hypothesise that particular language structures allow the production and reproduction of ideology, most of the time without the recipients of discourse being aware of it. Even if this an idea we do not find explicitly spelled out in many publications, CL usually considers language to be inherently ideological. As such, it can be exploited by dominant groups in an attempt to control other social groups. On the more technical side, CL draws on Halliday's systemic-functional grammar, which we will discuss further below. However, it should already be noted that such a linguistic take establishes one major difference between post-structuralist studies and CDA, namely the latter's focus on linguistic strategies in the analysis of social inequality (cf. Hart 2010).

Though CDA encompasses a very comprehensive and heterogeneous body of research, we will focus here only on those approaches which advocate an analytically prominent linguistic take. These are the mainstream approaches to which we will refer as CDA in what follows.

The linguistic component of CL, and to some extent also of CDA, is drawn from Systemic Functional Linguistics (henceforth SFL); the origins of this theory of language actually trace back to J. R. Firth's work in the first half of the 20th century. However, the main developments of SFL are attributed to M. A. K. Halliday, Firth's student, who elaborated these ideas into a full-fledged linguistic theory from the 1960s on.

One of the core tenets of SFL is that language fulfils specific functions through communication. People use language to achieve specific purposes, and these purposes in turn explain why language is the way it is. Instead of starting from the conceptual and abstract structure of language (*contra* e.g. Saussure) and thereby avoiding an abstract and exclusively formal or syntactic perspective, SFL leans more towards the study of what Saussure called the *parole* (the actual and variable realisation, in speech, of an abstract system of signs) by considering what linguistic-external functions language can serve. The unit of analysis is thus not the sentence, but the text, for the notion of usage is believed to be more relevant to functionality than that of grammaticality. Consequently, broader considerations of linguistic context and actual language use turn out to be crucial features of description. This echoes Firth's famous quote: "You shall know a word by the company it keeps" (1957: 11).

So SFL advocates that language cannot be accounted for in isolation, as an abstract system of meaningful units. The grammatical system is believed to be externally determined by socio-cultural factors:

> "Why is language as it is? The nature of language is closely related to the demands that we make on it, the functions it has to serve. In the most concrete terms, these functions are specific to a culture…. The particular form taken by the grammatical system of language is closely related to the social and personal needs that language is required to serve." (Halliday 1970: 173, quoted in Fowler 1994: 32)

It appears from this that the very nature of language is taken to be determined by factors which lie outside language itself; the functions language is said to fulfil constrain language and, by extension, language use.[73] Halliday postulated that language performs three main functions. The *ideational function* is the fact that language is a referring system which allows its users to embody their experience of the world into it, in order to express and communicate specific contents (i.e. meaning). The *interpersonal function* of language is its ability to relate language users to other communicators, thereby allowing for social interaction around specific issues to develop and be maintained; this function also supposes that language allows communicators to exchange attitudes, beliefs, evaluations, etc., about these issues, and could be summarised very simply as the possibility to interact with other people using a common means (it is thus a social function of language). The *textual function* focuses on the message, and stipulates that meanings can be encoded and decoded in an adequate way in language; it is a function which is centred on language itself.

Critical linguists, who were interested in studying how control could be achieved through language, found in SFL a powerful tool to perform linguistic analysis. Fowler acknowledges in this respect that "the best model for examining the connexions between linguistic structure and social values is the functional model developed by M. A. K. Halliday and his colleagues" (Fowler 1991: 68). Since under these assumptions language is as it is because it serves specific purposes, notably social ones, the particular way language is structured must somehow map the text producer's vision of the world and the purposes she wishes to fulfil in communication (in ideational and interpersonal terms foremost). In addition to echoing Sapir-Whorfian assumptions and ruling out the arbitrariness of language structure, SFL constitutes a technical way of analysing text according to social constructionist principles: "the forms of language encode a socially constructed representation of the world" says Fowler about the underlying assumptions of SFL (Fowler 1991: 37). In other words, the underlying structure of language is a direct reflection of the speaker's particular way of conceiving the outside world.

---

[73] Though it should also be noted that scholars drawing on SFL also usually consider that language use itself ends up constraining language; I will come back to this idea in the next section, as I discuss the relationship CDA establishes between discourse and social practice.

From a technical point of view, one of the ways of going beyond structure in order to assess how meanings are represented and organised in language (as part of its ideational function) is to look at *transitivity*. Unlike traditional grammars, SFL conceives of transitivity in semantic terms, as a sort of action-related template whereby agents do things to patients in particular ways.[74] As such, the representation of an event or a situation may be realised in different ways by text producers; the linguistic system indeed affords a plurality of combinations to linguistically encode the same event (the possibility of opting for the passive vs. active constructions is one example). The rationale for looking at how events are talked about in text and talk, in representational – or ideational – terms draws on the idea that different representations of the same event or situation convey different meanings (or make more salient different elements of the utterance). Since the way events are represented is believed to fulfil a purpose, it follows that the choice of a particular structure is never gratuitous, and that it always gives hints about the text producer's own representation of the world: "Since transitivity makes options available, we are always suppressing some possibilities, so the choice we make – better, the choice made by the discourse – indicates our point of view, is ideologically significant"[75] (Fowler 1991: 71).

II.1.1.2. On the scientificity of CDA: theory, multidisciplinarity and empirical research

II.1.1.2.1. The status of CDA as a scientific theory

CDA practitioners do not usually claim to contribute to a full-fledged theory of communication; rather, they consider theirs as an interdisciplinary programme drawing from various theories and approaches which are concerned with the nature and function of communication in society. As Wodak (2006) and van Dijk (2008) respectively formulate it:

> "It is necessary to stress that CDA has never been and has never attempted to be or to provide one single or specific theory. Neither is one specific methodology characteristic of research in CDA."[76] (Wodak 2006: 2)

[74] This is a core aspect of SFL-inspired linguistic analysis, which CDA significantly draws upon. The 'who-does-what-to-whom-and-how' scheme is taken as a structure which reveals the text producer's representation of events and situations.

[75] Notice here the somewhat odd role given to discourse: according to Fowler, speakers do not make choices; it is discourse itself which makes the choices that are found in actual utterances.

[76] In the remainder of the paragraph from which this quote is taken, Wodak stresses that any criticism of CDA (as a theory), cannot hold due to the heterogeneous nature of the research programme, and therefore that criticism should specify to which research or which researcher it is addressed. This *a priori* disqualifies any criticism against CDA as a whole. However, as we will see further on, all CDA approaches

> "CDA is not a theory or a method of research, but an academic movement of scholars specifically interested in the analysis of fundamental social problems, such as the discursive reproduction of illegitimate domination." (van Dijk 2008c: 821-822)

Following these considerations, it appears that CDA is not a single theory but an approach to communication that takes input from various theories, devoid of the theoretical ambitions other approaches to communication may claim to fulfil. As a consequence, CDA does not position itself as a unified scientific paradigm (understood in line with the criteria philosophy of science proposed, in the vein of Popper's (1959) and Lakatos'(1978)[77]). It does make use of findings resulting from extensive and sometimes extremely detailed scientific analyses, but its primary purpose is socio-political, rather than theoretical or scientific. This probably explains why, in Chilton's terms,

> "CDA, in its later manifestation has made, so far as one can see, no contribution to our scientific understanding of the language capacity in *homo sapiens*." (Chilton 2005a: 22)

It is a natural consequence of CDA's inherent tension between the scientific and the political that different types of research practices emerge, because relaxing the constraints on the type of scientific practice to be adopted affords a plurality of directions of research to be explored; nowadays, there are accordingly roughly three orientations in mainstream CDA: Fairclough's macro-analysis of discourse and social change (cf. Fairclough 2001 [1989], 1995, 1992), Wodak's discourse-historical approach (cf. Wodak in Wodak & Meyer 2001, chapter 4), and van Dijk's socio-cognitive model which focuses on the cognitive aspects of discourse processing (cf. van Dijk 2008a and 2008b).

Looking at things from an empirical point of view, the endorsed multidisciplinarity outlined above unsurprisingly results in an important collection of case studies. CDA is about concrete analysis, and thus requires the study of naturally occurring data. Since the core assumptions of CDA require that the issue to be addressed relates to social configurations of inequality and power/domination in interaction, any social communicative situation displaying such characteristics can be accordingly analysed as corpus. That is why mainstream CDA research focuses to a great extent on discrimination and social asymmetry (e.g. social, racial, gender-motivated, etc.) usually by examining the way communities and social practices are represented in text and talk. Accordingly, analysis of real data is a key element of CDA, and recourse to

---

do share certain assumptions about what discourse is and how it functions, which does allow some discussion and criticism.

[77] See Chalmers (1982) for an informed discussion of these approaches.

corpora constitutes the empirical warrant of the analysis. It is also for this reason that CDA so strongly emphasises methodological issues of both collection and analysis of real data alongside the need for corpus analysis (see e.g. Wodak & Meyer 2001). It is equally revealing in this respect that some trends in CDA have recently begun to integrate corpus linguistics research for the analytical benefits the latter may yield *vis-à-vis* its systematic handling of corpora (cf. Baker *et al.* 2008, Mautner 2005, 2008). Consequently, a great many CDA practitioners tend to highlight the practical analysis of social inequality, and as a consequence sometimes pay less attention to the epistemological and theoretical aspects of their own scientific enterprise. It is however clear that the gist of CDA rests in its analytical instrumentality rather than on an ambition to develop a full-fledged theory of discourse.

Yet, a few strikingly relevant quotes taken from mainstream CDA reveal in this regard an ambiguous position on CDA's scientific status as far as theory is concerned:

> "CDA is informed by social theory and views discursive and linguistic data as a social practice, both reflecting and producing ideologies in society (of course, all scientific endeavour is socially committed, as Habermas, 1967, clearly illustrated for the social and natural sciences). In this way, *all CDA approaches have to be regarded not only as 'tools', but also as discourse theories*." (Baker *et al.* 2008: 280, my italics)

Here Baker *et al*. state i) that they construe their work as an analysis of discursive social practices and of the way these may (re)produce ideologies, but also ii) that their endeavour bears a theoretical status. The problem in this quote is to establish the link between i) and ii), i.e., to delimit the scope of theory within CDA. On the one hand, one could consider that Baker *et al*.'s practice claims theoretical status by virtue of its making use of social theory – but also of theories of communication in the broad sense, were we to consider 'discourse' from the perspective of the production and reception of communication –, which is indeed something very common in CDA, mainly due to its multidisciplinary perspective (see. §II.1.1.2.2.1. below). CDA draws on theoretical claims from other disciplines in order to sustain its analysis; the theoretical nature of CDA, under this view, is thus not exactly CDA's own, since it is to be found in the various theories it borrows from, and this is quite compatible with Wodak's and van Dijk's quotes reproduced above.

Now, if we consider that the expression 'discourse theory' is taken to scope over CDA as a discipline in its own right, independently of any theory it may draw from, then we need to accept that CDA indeed bears theoretical ambitions, in the sense that it develops an account meant to generate claims to knowledge and that the latter is designed to explain the behaviour of some object in the world by uncovering cause to effect relationships. In other words, there are reasons to consider that the term 'theory' in CDA can at times scope over its own practice; yet, and

perhaps paradoxically so, given CDA practitioners' claims to avoid this, it would force us to accept that CDA nourishes the same ambitions any full-blown scientific theory nourishes.

A simpler way to formulate the question at this point is to ask whether CDA denotes a set of various discourse/communication theories, or if it can be taken as one single theoretical paradigm. In Wodak's (2006) terms, quoted above, it seems that we should prefer the former option: CDA is not a single theory, but a set of approaches sharing a common agenda. Whether this agenda can be assigned a clear theoretical status remains doubtful for the moment.

It might come handy to consider what other prominent figures of CDA have to say about this. Consider in this respect the following quote from Fairclough: "[…] in the longer term critical discourse analysis should, as Kress has argued (1993), be informing the development *of a new social theory of language which may include a new grammatical theory*" (Fairclough 1995:10, my italics). Here Fairclough explicitly speaks of a *single* "social theory of language", and clearly states that the elaboration of such a theory, which includes a linguistic component – "grammatical" in Fairclough's terms probably points here to a development of Hallidayan systemic-functional linguistics –, should be achieved *via* critical discourse analysis. In other words, Fairclough points to the fact that CDA (and we have no clear indication here that we should not interpret CDA as referring to a single theoretical field) is indeed a scientific enterprise, and that as such, it should bear theoretical status, as it should contribute to the elaboration of a theory.

So, on the one hand, Wodak and van Dijk argue that CDA is not a theory; on the other, Fairclough leaves the door open to the converse possibility. If we are to consider Fairclough seriously, then we need to identify the epistemological nature and features of CDA as a unified theory. However, as Wodak and van Dijk claim, such a unified epistemological set of principles can never be found in CDA, which is multidisciplinary by nature. In that case, what makes CDA CDA is not a shared set of theoretical principles, but rather a shared *agenda* which stems from the very meaning of the word 'critical' and which carries no obvious epistemological requirements, since a research agenda in this vein escapes the constraints traditionally attached to a scientific theory.

Let us now consider Kress' quote to which Fairclough referred previously; Kress also explicitly formulates that any critical study must develop a theory of language, and, by extension, of communication: "[…] it has become essential to take a decisive step towards the articulation of the theory of language, or communication, of semiosis, which is implied in these critical language activities, *to develop an apt theory of language*" (Kress 1993: 15, my italics).

Whether CDA as a whole is a proper full-blown theory or not remains unclear to me; it turns out that CDA practitioners don't seem to agree on the matter themselves. Even if elsewhere she

argued otherwise, there are still reasons to suspect that Wodak endorses the theoretical status of CDA: "In fact, in recent research it seems that the term CDA is preferred and is used *to denote the theory* formerly identified as [Critcal Linguistics]" (Wodak 2007: 208-209, my italics).[78] So on the one hand we get the impression that CDA denotes a collection of independent approaches making use of some other theories, while on the other sometimes we are left with a more unified perspective.

The relevant question to ask is thus the following: would the latter theoretical requirement have to apply indiscriminately to all CDA studies – or 'approaches' – or would we need to consider each CDA approach as a theory, just like the plural in "discourse theories" (Baker *et al* 2008: 280) seems to indicate? The question is far from being clear, and, to my knowledge, there is little discussion about these epistemological issues in CDA.[79] In light of the positions reported above, I will thus consider the scientific/theoretical status of CDA to be problematic; it should be nonetheless noted that what Wodak and van Dijk argue in favour of – i.e., a 'non-theoretical' status – seems to be more in accordance with the type of multidisciplinary work we actually find in CDA studies. Moreover, it is not unusual to find, among CDA scholars, some kind of suspicion towards science; scientific practice, taken as a discursive practice (or as a discourse), is believed to be susceptible to ideological bias just as any other discourse, and thus it must also, and probably above all discourses, be submitted to critical scrutiny. The problem being that sometimes scholars defending that viewpoint end up disqualifying scientific investigation on the grounds that it would be a partial and biased expression of an ideological stance, without engaging in an assessment of whether the research is sound and relevant according to other, perhaps more objective, criteria.

II.1.1.2.2. The tension between the scientific and the political

II.1.1.2.2.1. A criterion of utility for multidisciplinary selection

As Wodak (2006) and van Dijk (2008a) note that the common feature of all CDA research cannot be found in a unified epistemology nor in a pooled methodology – since one cannot identify a specific set of foundational principles and methods underlying *all* CDA work – common features of CDA research are defined through a shared agenda (and, additionally, I would venture, through shared conceptions about language and discourse).

---

[78] In all fairness, it should be added that in this sentence, Wodak is merely reporting a widespread position on what CDA denotes.

[79] I would tentatively venture that the difficulty in resolving this issue stems from a tension between the scientific status of CDA, its problematic borrowing from other theoretical frames, and its avowed activist stance (see §II.1.1.2.2. below).

What unifies CDA approaches is thus firstly a common critical stance; a major claim one can make to characterise all CDA studies is that these are motivated by a social purpose, in an activist sense. Analysis should serve an emancipatory cause, so as to "root out a particular kind of delusion" (Wodak 2006: 3). Hence, CDA is purposeful from a social point of view. In turn, scientific requirements for the construction of a unified theory are not regarded as elements of paramount importance.

Relaxing epistemological constraints encourages the integration of numerous trends in linguistic and communication research. Multidisciplinarity in CDA lies in the diversity of methodologies and theoretical approaches employed; over the years, recourse to fields of research as varied as Conversational Analysis, Corpus Linguistics, Content Analysis, Discourse Analysis of several traditions, Ethnography of Communication, Systemic-Functional Grammar, Text Linguistics, Rhetoric, Argumentation Theory, Pragmatics, Cognitive Linguistics, Frame Semantics, Cognitive (or Mental) Models Theory, Script Theory, Semiotics, Social Semiotics, among others, has contributed to develop and enrich research in CDA. Indeed, as CDA is not driven by a unified theoretical goal (such as providing a theoretical model of human communication), but rather by the prime concern about "the discourse dimensions of power abuse and the injustice and inequality that result from it" (van Dijk 1993: 252), it follows that the theoretical apparatus it employs is "chosen or elaborated as a function of [its] relevance for the realization of such a socio-political goal" (*ibid*.). Consequently, CDA has borrowed from all these disciplines an array of analytical and conceptual instruments which it has used to deconstruct the way ideology functions – most of the times covertly – in discourse. The result of such an integrative endeavour is the elaboration of a multifarious and heterogeneous toolkit of linguistic analysis.

The criterion according to which relevant tools are selected is a practical one: it is by virtue of their linguistic-analytic *usefulness* that CDA practitioners judge insights from other disciplines to be relevant and applicable to their own research. Wodak (2006: 82) recalls Mouzelis when she advocates a "conceptual pragmatism" which privileges utility over truth, subordinating at the same time the importance of the epistemological features of the account to its social usefulness in fulfilling emancipatory goals. This in turn also explains why van Dijk considers that in CDA "the distinctions between theory, description and 'application' become less relevant" (van Dijk 1993: 252). In other words, the epistemological properties of all these input theories do not receive as much attention in CDA as their analytical usefulness, the risk being free and intuitive speculation.

As an academic discipline meant to serve a social purpose, and despite researchers claiming otherwise, CDA is characterised both by a scientific and a political agenda. Widdowson (1995, 1996, 1998, 2000) suggests that this tension can be problematic in many respects, and I will now

turn to discuss some of his arguments in order to deal with questions related to the multidisciplinary practice CDA engages in.

From the point of view of the epistemological feasibility of such a massive integration, three important questions arise: i) can we justify in theoretical and epistemological terms the use of such a heterogeneous pool of theoretical input into a single paradigm?, ii) if not, does this threaten CDA's own epistemological credibility as an enterprise of scientific inquiry, and, in this case, iii) can we salvage CDA's scientific status? These questions are seldom addressed in mainstream CDA;[80] they nevertheless deserve to be answered if we are to evaluate the consistency of the scientific practice itself.

II.1.1.2.2.2. Scientificity and activism: an oscillation between constraints

In a series of papers reviewing work in CDA from a critical perspective Widdowson takes the tension between scientificity and political agenda to be one of the main drawbacks of CDA. Though some of his radical statements can to some extent be taken as overly (and even perhaps unnecessarily) evaluative, some of the criticisms he addresses to CDA do hold, and they contribute to shedding light on the epistemological issues I have evoked in the formulation of the three questions above (see §II.1.1.2.2.1.).

One of the most provoking conclusions he reaches through his review lies in his idea that, by definition, CDA cannot be analysis, and that the output of this type of research is necessarily interpretative: "if critical discourse analysis is an exercise in interpretation, it is invalid as analysis. The name 'critical discourse analysis', in other words, is, in my view, a contradiction in terms" (Widdowson 1995: 160).

The grounds on which he makes such a claim are to a large extent derived from empirical evidence. In his (1996) "Reply to Fairclough", he takes some fragments of corpus Fairclough analysed himself in his (1992) book, and goes on to show that the analysis the latter provides can be seriously challenged by a series of alternative interpretations of the same corpus. In particular, Widdowson analyses some of the linguistic items and devices found in that corpus to show that they can also be taken to lead to opposite interpretive conclusions.[81] Obviously, providing counter-arguments on the basis of the same kind of analysis constitutes a major problem for the validity of the original claims, and in Widdowson's opinion, this result points to

[80] An exception to this is Weiss & Wodak (2001), where the first part of the chapter addresses the epistemological foundations of CDA.

[81] See Widdowson (1996: 61-69) for the details of the analysis.

a lack of rigour in Fairclough's work: it shows that his analysis does not adequately resist objections. Widdowson then ventures that this can be explained by the fact that in his opinion the CDA analyst "preconceives the data as evidence. Fairclough knows what he is looking for. Not surprisingly, he finds it" (Widdowson 1996: 63).

In a publication similar in topic, Widdowson takes as an example an analysis of a press article of the *Sun* by van Dijk (1996), and goes on to show that in the same paragraphs where van Dijk finds traces of racist ideology, we can also find plausible traces of colonial conscience, thus revealing a less racist attitude than what van Dijk finds in the text. That is, the same excerpt can be taken to contain linguistic evidence of contradictory attitudes.[82] So, in a way, as Widdowson notes, it is as if

> "van Dijk is looking only for textual confirmation of a bias he has attributed to the source of his text in advance. Everything that appears in the *Sun* is necessarily racist. So in effect, what van Dijk is doing here is controlling our access to this text by imposing his own discourse upon it." (Widdowson 1998: 144)

Beyond the obvious problems such examples can pose to CDA's credibility as a consistent scientific practice, what is perhaps more worrying is that, while mainstream CDA is perfectly in agreement with the idea that different complex meanings may be read off a text, therefore acknowledging that interpretation is more a pragmatic than a semantic matter, it does not propose any theoretical apparatus meant to explain how these different meanings can be retrieved. We are left with the observation that (parts of) a single text can be interpreted in different ways, but also with the impression that CDA does not give any justification other than the analyst's intuition as to how we should favour one over another; and, in the end, since the interpretation one reaches depends on what one reads into the text, there is no guaranteed criterion to ascertain whether a given text is ideologically loaded or not, which is highly problematic for the point CDA wants to make.

The reason for this, Widdowson claims, is that CDA practitioners, in their analysis, already know what they are looking for, for they start off with an activist stance geared towards the identification of illegitimate or problematic ideological manoeuvres in text and talk. As a consequence, he claims that their analysis is already *a priori* biased and that their practice is merely confirmative: it is meant to illustrate that such and such corpus is ideologically loaded, which is the starting point of the analysis. In other words, the analysis is prejudiced even circular, since it starts with an identified ideological claim in order to validate it empirically.

[82] See Widdowson (1998: 143-144) for the detailed example.

Now, from a scientific perspective, this, according to Widdowson, is interpretation, not analysis, because proper analysis cannot afford political commitment, (thereby grounding his conception of scientific practice on an objectivist stance). In his view, analysis should make explicit the different interpretations a text can receive, and moreover the mechanisms by which those interpretations can be retrieved in order to sustain the particular reading to be found in the text, while at the same time refraining from providing a single interpretation presented as the correct one. But he highlights that we get nothing of this sort in CDA; all we get is *one* interpretation of the data (among possible alternative interpretations) which is furthermore not scientifically justified.

Of course, this does not invalidate the particular interpretations arrived at *per se*, and nor does it invalidate altogether the point of social significance the study wants to make, but it does entail that *one interpretation among others* is provided, which, in principle, should not be more valid than an alternative one, unless the study provides conclusive evidence to favour it. Furthermore, Widdowson suggests that CDA leans more towards social significance than towards linguistic analysis. He also points out (Widdowson 1995) that as CDA practitioners are aware that interpretations of discourses may vary from one person to another (van Dijk explains this in terms of context models), they, above all people, should be aware of their own ideological bias when performing analysis, or, as Widdowson calls it, interpretation.

In sum, Widdowson files a charge on partiality on two counts: "first, [CDA] is not impartial in that it is ideologically committed, and so prejudiced; and it is partial in that it selects those features of the text which support its preferred interpretation" (Widdowson 1995: 170).

The purpose of this section has been to illustrate that scientific concerns, in CDA, are sometimes subordinated to ideological/activist purposes. This may turn out to be problematic because it makes CDA practice almost exclusively interpretative; there is nothing wrong with this from a social perspective, but from a scientific perspective it could hinder CDA's credibility as a scientific endeavour, since the criteria by which we may scientifically ground its validity – i.e., by providing sufficient reasons to consider that the interpretation provided is indeed valid – do not resist empirical falsification. With this an evident risk emerges, which Widdowson summarises in a (perhaps too) radical way:

> "The adherents of CDA, it might be suggested, are too anxious to make a political point, too quick to come to critical conclusions, and too ready to see themselves as engaging in a revolutionary kind of linguistics, one whose subjectivity (to use Kress' phrase) has certain 'orientations to 'rationality'', whereby it can set its own agenda and define its own conditions of validity. The agenda is the exposure of prejudice and the abuse of power. The conditions are commitment to an egalitarian cause. […] One can accept the cause as just.

> But the assumption seems to be that if the cause is just, this in itself validates the means of furthering it." (Widdowson 1998: 149-150)

The point is that if the researcher relaxes the constraint on scientificity, her/his account may end up losing strength, thus making the validity of the proposed account rest on the social significance of her/his analysis, so that moral considerations may suffice to convince the reader. As Widdowson puts it: "[CDA] carries conviction because it espouses just causes, and this is disarming, of course: it conditions the hearer into acceptance. If you can persuade people by an appeal to moral conscience, you do not *need* good arguments" (Widdowson 1998: 150, author's italics).

While one may not necessarily agree with what we might read into Widdowson's arguments as far as the integrity of CDA is concerned, it seems fair follow him on the risks attached to the excessive prominence of the analysis' social goal, especially since the scientific merits of the discipline become consequently questionable.

In what follows I question further epistemological implications of multidisciplinarity which, in my view, also relate to CDA's inherent oscillation between science and social commitment.

II.1.1.2.2.3. Epistemological issues associated to multidisciplinarity

As mentioned above, the body of research CDA draws from is considerable. One could wonder if such a massive and heterogeneous integration scientifically goes without raising any issues beyond a criterion of utility, that is, on the grounds of epistemological considerations. Let us take an example to illustrate the point.

Contemporary Pragmatics (in particular post-Gricean accounts), which CDA draws on for instance when it makes use of the notion of implicature[83] (following Grice's [1975] 1989 seminal account), is a field of research which typically takes intentionality to be one of the key features of communication by postulating the underdeterminacy of meaning (whereby we communicate much more than what is contained in the linguistic code); in ordinary conversation participants are accordingly believed to make assumptions about each other's intentions, beliefs, desires – i.e., about each other's mental states – in order to retrieve the meaning of communicative utterances. This is underlain by an epistemological stance which takes the human mind to be an object of scientific investigation in its own right. And indeed, the majority of pragmatic approaches to communication, after Grice, took on the development of cognitive models of information-processing (e.g. Horn 1984, Bach & Harnish 1979, Levinson 1983, 2000, Sperber &

[83] See for instance van Dijk (2005a).

Wilson 1995, and Carston 2002a). The field of experimental pragmatics has also taken this assumption seriously and has relatively recently started to test some of the pragmatic principles postulated in theoretical work (see Noveck 2001, Noveck & Sperber 2004).

It would be fairly sensible to consider that when CDA borrows tools from Pragmatics it should also take on board the theoretical underpinnings associated to the concepts which are used in that discipline. It would indeed be a paradox to use the notion of implicature without considering i) that speakers have intentions, ii) that they want to fulfil these intentions through conversation, and iii) that addressees are able to recognise such intentions on the basis of the speaker's contextualised utterance. Furthermore, pragmatic research usually adopts a (hypothetico-) deductivist epistemology, whereby the researcher is granted the possibility of making hypotheses and testing them according to different standards, one of them being introspection. Introspection is accordingly considered in this paradigm[84] to be an acceptable way of reaching claims to knowledge about language use, since we are all deemed competent pragmatic users of our native languages. It is reasonable to assume somehow normatively that a fair, informed and adequate borrowing of pragmatic conceptual and analytical notions should not be dissociated from the epistemological implications attached to them. Otherwise, it is expected from the borrower that s/he make explicit and explain why s/he disposes of these.

I have noted that CDA chooses its analytical tools from a large array of approaches to communication. Recourse to Conversational Analysis (following the works of Harvey Sacks, Emanuel Schegloff and Gail Jefferson, henceforth CA) despite having given rise to scientific debate between proponents of both disciplines[85] is often put into practice in CDA, notably in what refers to data analysis of naturally occurring, socially situated, text or talk. Van Dijk (1999) notes in this respect that the two approaches share a number of assumptions, both in their methodology and epistemology (which by the way suggests that van Dijk does in the end consider that his research bears epistemological requirements, which in turn highlights that, despite all claims, he does consider his work to be theoretically grounded).

CA seeks to describe and explain the order of communicative interactions; as such, its tone is sociologically-driven to a large extent:

> "CA is only marginally interested in language as such; its actual object of study is the interactional organization of social activities. […] the production of utterances is seen […] first and foremost as a practical social accomplishment" (Hutchby & Wooffitt 1998: 14).

[84] Following the line of thought defended by mainstream cognitive science in the Humanities.

[85] See for instance the exchange between Billig and Schegloff in issue 10 (4) (1999) of *Discourse & Society*.

We can note that a first difference between CA and standard contemporary Pragmatics lies in the former's conception of the production of utterances being seen as a social accomplishment, whereas the latter focuses on the cognitive dimension of communication. Though I do not believe that these opposite perspectives cannot be reconciled,[86] they often give rise to strong antagonisms among schools of thoughts.

Focus in CA is set on data, and on what data can tell us about the ways interactions are structured. As such, it avoids theoretical preconceptions in order to prevent the researcher's bias from entering the analysis, and also tries to avoid recourse to conversation-external information, in a 'what-you-see-is-what-you-get' fashion. The epistemology defended in that framework is thus presumably inductivist, albeit with an objectivist tone, whereby it is considered that the explanatory account must emerge from the data, and neither from "introspection" (i.e., from empirical data about linguistic and pragmatic competence) nor from previous theoretical assumptions, which is a vision quite different from what we find in contemporary Pragmatics. Furthermore, CA does not take intentionality to be a proper object of inquiry, since it shares with behaviourist approaches the idea that we can only study the actual observable behaviour of subjects in interactions; accordingly, conversational analysts very often doubt that a proper study of what happens in the mind can be achieved, unlike the study of text or talk, for its material, observable, existence cannot be accessed. Finally, a last difference between CA and Pragmatics worth mentioning stems from the data-driven requirement of the former: it would seem impossible in principle to host any account of implicature in CA, since its derivation precisely requires conversation-external, i.e. contextual, information.

Now, given CDA's commitment to multidisciplinarity, which in itself, I believe, is a necessary component of an informed account of communication, one could question the coexistence of approaches as different as CA and Pragmatics into a same research programme. If borrowing tools from one discipline also entails acknowledging and inevitably supporting its epistemological underpinnings, then CDA, on the scientific level, faces a problem when it uses concepts and tools from epistemologically incompatible approaches, such as the two mentioned above.

A first answer one could formulate to defend multidisciplinarity in light of the above objection would be to consider that it is not problematic to make use of epistemologically incompatible approaches as long as i) we do not use their tools in the same analysis of data or in the same area of the account, or ii) the piece of research in question makes use of only one methodology, so as

[86] See for instance Clément (2006) for a model of belief fixation informed both by cognitive and social accounts.

to maintain epistemological consistency. Converging with Wodak's (2006) claim whereby CDA cannot be criticised as a whole, unified, theory, we would posit that what defines a piece of research as CDA is its critical attitude, thus accepting that there are variable types of CDA which can coexist in the programme: Pragmatic CDA, CA CDA, Corpus Linguistics CDA, Cognitive Linguistics CDA, and so on. This in itself is not a problem, but for the sake of theoretical consistency, it does require from each of these studies that they explicitly opt for one particular epistemology, and that they refrain from adopting epistemologically incompatible tools. So the first answer to the problem of multidisciplinarity would be to break down research in CDA into specific fields whose analytical practice is defined in terms of the discipline it borrows its tools from. This would certainly warrant the conception of CDA as a collection of approaches to discourse, communication and society, devoid of theoretical unity. Under this view, each CDA study in itself is a unique piece of research, and as such it would be up to each particular researcher to choose the theoretical tools s/he judges adequate with respect to the point of the study.

An alternative would be for CDA research to discuss, in epistemological terms, how different approaches may be interfaced, taking into account the resulting philosophical implications for the scientific practice. But this is not the purpose of CDA; social significance and practicality of the analysis primes over other concerns, which is the reason why such pieces of research are seldom found within the research programme.

While this seems to be the preferred solution to the problem of multidisciplinarity, positions on the theoretical status of CDA such as Fairclough's, Kress' and Baker *et al*'s remain problematic, because there is an inevitable trade-off if you opt at the same time for fragmented multidisciplinarity and theoretical unity. The criterion of utility is crucially detrimental to theoretical consistency, for the coexistence of incompatible approaches into the same model cannot be resolved unless a detailed discussion is undertaken to reconcile divergent aspects of the latter. This would mean that these ambitions should be abandoned in order to achieve unity, that is, unity in a social agenda, not in a theoretical frame.

The conclusion we can draw from these considerations relating to the problematic integration of epistemologically incompatible scientific traditions, is that, if taken as a systematic scientific enterprise, CDA faces serious problems. Solving these would require either a thorough discussion of philosophical and epistemological issues, which CDA does not typically engage in, or the dismissal of such worries on the grounds of the criterion of utility advocating conceptual pragmatism, which, at the end of the day, seems to be the preferred solution (if we follow van Dijk and Wodak). What I am suggesting is that the reconciliation of science and social activism is not self-evident; in the ideal world of epistemology, we would have a clear-cut stance, on behalf

of the researchers, about the scientific foundations of every piece of research. CDA offers a multiple stance, which can vary from one study to the other, but which does not solve the issues expounded above.

So, in order to answer the three questions raised above, let me briefly recall them and sum up what has so far been discussed:

> (i) Can we justify in theoretical and epistemological terms the use of such a heterogeneous pool of theoretical input into a single paradigm?
>
> With difficulty. From an epistemological perspective, it would seem quite implausible to justify the coexistence of epistemologically divergent analytical and conceptual tools into the same model. This is due not to CDA itself, but to the incompatibility of approaches CDA borrows from.
>
> (ii) Does this threaten CDA's own epistemological underpinnings?
>
> It depends. If we take seriously the theoretical ambitions formulated by Fairclough (1995), Kress (1993), Baker *et al* (2008) and to some extent in van Dijk's (1999) claim (about CA and CDA sharing epistemological features), then yes, CDA is indeed threatened, because of the reason evoked in i) above. However, if we shift the perspective to social usefulness and stick to critique, leaving scientificity aside, then the answer is no; there is no threat, because there are no epistemological requirements to be threatened in the first place. But this is problematic for the following point.
>
> (iii) Can we salvage CDA's scientific status?
>
> This is a highly problematic question, since it presupposes that CDA does have a scientific status, which is something CDA practitioners do not agree themselves on (in terms of the satisfaction of a stable set of epistemological standards); in other words, it amounts to questioning the very idea that CDA is a theory. If we go by the widespread view that it is an approach rather than a theory, then, again, the question does not apply. Yet, it should be added that even if CDA is not a theory, all CDA approaches share certain basic philosophical assumptions (such as, for instance, the conception of discourse as a social practice or claims about the ideological load of language, which it inherits from its historical roots).[87] So even if those do not constitute a definite core set of epistemological claims designed to

[87] See §II.1.1.1.

> underlie a consistent theory, they still bear epistemological implications. Thus I take it that the question of salvaging CDA's scientific status, in this respect, can only be answered through an assessment of these shared claims. This will be the purpose of §II.1.1.3.

My purpose in this section was to assess whether research in CDA could be an epistemologically interesting option for the type of account I am out for; I acknowledge that I am more inclined to assess scientific rigour and validity than social significance, though I believe that the former may make the latter stronger.[88] This is why my perspective seems to depart from CDA's. It is my intention to try to assess whether the line of work pursued in mainstream CDA can shed light on the issue of uncooperative and manipulative communication; I have tried to show so far that the fragmented epistemological nature of this particular approach to communication makes it very difficult to evaluate it as a scientific practice and virtually inoperable with regard to the scientific goal of explaining non-cooperative and manipulative communication.

I now turn to discuss additional, practical, risks associated with the tension CDA faces between scientificity and social activism, for I believe that much of the confusion Widdowson identified within CDA stem from this twofold problematic endeavour.

II.1.1.2.3. Further risks of instrumental multidisciplinarity: misuse of conceptual tools and over-interpretation

An additional risk attached to an instrumentally-(but not epistemologically) motivated multidisciplinarity relates to a debate-laden concern Widdowson (1995a, 1995b, 1996, 1998) and Billig (2008) have voiced and which boils down to the idea that relaxing the epistemological constraints of multidisciplinarity for the sake of usefulness carries the risk of misusing the analytical and conceptual tools one borrows from other disciplines. In what follows I will argue, by taking three particular examples, that the promptness to make a political point can indeed sometimes lead researchers in CDA to unwittingly misuse conceptual tools which have traditionally been developed in other paradigms. I will take this issue to be a natural consequence of the type of "selected multidisciplinarity" (cf. Chilton 2005a) advocated by CDA practitioners.

---

[88] Science should be at the service of 'higher' causes: I believe that a rigorous and strong explanatory account can only strengthen the social significance of a point: in particular, regarding phenomena of manipulative communication, I contend that knowing how manipulation works will make you more apt to resist it.

II.1.1.2.3.1. The case of van Dijk's "political implicatures"

In a paper dealing with the political function of implicit components of meaning published in 2005, van Dijk defines a new category of implicature, the so-called 'political implicatures'. The paper suggests that by taking a very close look at these implicit contents, analysis can gain descriptive value and thus contribute to a more profound understanding of how political goals, such as positive Self representation and negative Other representation for instance, may be attained. I contend that van Dijk's resorting to the notion of implicature to this end is problematic at least on two counts, and that as such it contributes to illustrating the risk associated to the type of multidisciplinarity usually advocated in CDA.

The first point of discussion concerns definitional issues and specifically terminological and categorical aspects. By naming a certain type of implicit contents 'political implicatures' without further discussion, van Dijk could be taken to suggest that these constitute an additional formal category of implicature, next to the traditional categories we find in pragmatic research (following Grice 1989), namely conventional and conversational implicatures (the latter being broken down in turn into generalized and particularized implicatures). This terminological choice could come as a surprise, since from a standard pragmatic perspective, it makes little sense to define an implicature in terms of its scope, function or topic – which is what van Dijk seems to do by using the term 'political implicature' – because an implicature denotes the outcome of a specific inference calculated on the basis of a communicative stimulus (sometimes combined to contextual information for conversational implicatures), independently of its topic, function in the exchange or thematic scope. In other words, implicatures are traditionally characterised and distinguished in terms of the way they are derived, not in terms of their topical or functional belongingness. If such were not the case, we could multiply categories *ad infinitum* so as to include in a pragmatic account items like 'capitalist implicatures', 'moral implicatures', 'religious implicatures', 'legitimising implicatures', 'apologising implicatures' 'love implicatures' and so on and so forth, which, as probably no pragmatist would deny, do not constitute relevant categories for the explanation of how language users derive implicit contents. The reason for this is that topical and functional belongingness are not decisive factors for implicature classification, as the implicature's propositional content and function in an exchange are by definition contextual variables which are independent from the derivation of the implicature *per se*.

The whole rationale behind Grice's original categorisation of implicatures into a finite set of categories lies in the postulates that i) these types of contents are not literally expressed, and ii) they can be derived in different ways, depending on the extent to which they involve more

contextual or linguistic factors. The difference between conventional and conversational implicatures lies in the former's reliance on the linguistic system and the latter's high dependence on the context of the utterance. Consequently, the criterion by which implicatures are traditionally distinguished has to do with the amount of linguistic *vs*. contextual information involved in their derivation, but neither with the content of the contextual assumptions that are mobilised in the process nor with the function the implicature may serve in a particular exchange.

Let us take a closer look at the way van Dijk construes 'political implicatures'. His account builds on the analysis of the opening sequence from a speech delivered before the Spanish Parliament by former Spanish Prime Minister José María Aznar about the government's position on the crisis with Iraq in 2003, reproduced in (34):

(34) "Señora Presidenta, señorías, al comenzar el periodo de sesiones, comparezco esta tarde ante la Cámara para informar a SS.SS de la posición del Gobierno ante la crísis que enfrenta la comunidad internacional con Irak" (quoted in van Dijk 2005a: 72)

[Mrs. President, Seigniories, as the session period begins, I stand this afternoon before the Chamber in order to inform you of the Government's position towards the crisis which the international community faces with Irak (my translation)].

Van Dijk defines political implicatures as context-sensitive contents, and more precisely contents partly derived from 'political' contextual assumptions, which can fulfil political goals in an exchange. This can be inferred from the way he characterises the sources of these implicatures (although he does so in a somewhat obscure fashion):

"Political implicatures are assigned by the participants as inferences from three sources:

i. participants' representations of the structures of the discourse and its meanings (such as their mental model of the situation in Iraq)

ii. participants' context model of the current communicative situation

iii. participants' more general knowledge about the political situation in the world and in the country." (van Dijk 2005a: 70)

Leaving aside that I find it hard to understand what it means exactly for a participant to "assign an implicature as an inference from a source", it appears that the list of three possible sources for implicature derivation drawn by van Dijk is little more than a specification of the nature and origin of contextual (in a broad sense) information a hearer may take into account in order to derive the implicature (when indeed there is one). This list stipulates that information about i) the meaning of the discourse, ii) context models of the situation and iii) background knowledge

are prone to serve as contextual assumptions in the implicature derivation process. Thus, beyond these distinctions van Dijk deems crucial, the contextual nature of these 'sources' suggest that what he has in mind is probably plain conversational implicatures. The relevance sought by naming them 'political' consequently appears to be a terminological convenience more than a theoretically motivated option, and should thus not be taken as such.

As for the choice of the term 'implicature', van Dijk motivates it as follows:

> "I have chosen the term 'implicature' rather than 'implication' because the inferences involved are not semantic, but pragmatic or contextual. (…) *Implicatures* on the other hand are usually defined as weak semantic implications or pragmatically in terms of contexts (Atlas 2000; Gazdar 1979; Grice 1989; Levinson 2000). My use of the term will here be limited to the pragmatics of context". (van Dijk 2005a: 69)

There is something oddly strange about this definition; as far as I am aware, no one in the literature on implicature has considered implicatures as "weak semantic implications". The inadequacy of the terms 'semantic' and 'implication' to denote implicit contents was precisely among Grice's motivations for coining the term *implicature*, meant to denote those non-truth-conditional aspects of meaning which escape the semantic content literally encoded but which are nonetheless communicated. Of course, one could suspect that it is not primarily the implicature's precise philosophical definition van Dijk is interested in, but rather the fact that it is contextually constrained. This can be deduced from the last part of the above quote, where he explicitly focuses on their contextual nature, which could in turn be a confirmation that what he really is after is conversational implicature. Still, such an odd characterisation (i.e. "weak semantic implications"), what is more erroneously presented as a mainstream conception, is false and its presence in the paper could probably be explained by van Dijk's eagerness to make a political point in the analysis, this in turn leading him to pay little attention to conceptual accuracy.

Now, suggesting that there is such a thing as a 'political implicature' would require a thorough justification that states precisely why and how the category 'political' is relevant next to the traditional categories (conventional, conversational), if only to be fair both to an uninformed readership with limited knowledge (or no knowledge at all) in standard Pragmatics and to the pool of researchers who have extensively dealt with the notion of implicature since Grice. But van Dijk does not make this explicit in his paper. Even if he states that his usage of the term 'political implicature' denotes a reduced set of contents in terms of their thematic or topical relevance in the exchange together with their 'political' function, this does not do justice to the pragmatic research that has focused on implicature over the past 40 years. We could therefore argue in principle that van Dijk neglects some basic and defining aspects of implicature by using

the notion of 'political implicature' as a relevant and appropriate concept, and that his account is at worst confusing, as it is not clearly positioned in light of previous research on implicature. When a researcher uses a notion borrowed from another field in a way that departs from its original use or defines it in a different way, we would normally expect these choices to be motivated.

In his paper, it appears that what he really is after is a development of the idea that non-literal contents may play a role in the propagation of ideology, the positive and negative representation of Self and Other, and more generally in discursive processes of persuasion. Yet, and this is something I will elaborate on further down, there are reasons to doubt that this is a goal that can be attained by resorting to the notion of implicature, at least under van Dijk's standards. I will argue that the notion of implicature is in fact unsuited to account for these possibilities if taken in van Dijk's perspective.

A second objection to van Dijk's account of 'political implicatures' has to do with his explanation of how these implicit contents are derived, which we find in a discussion of the example mentioned above (Aznar's speech). Taking van Dijk's more specific definition, political implicatures are meant to be conceived as "the specific political inferences that participants in the communicative situation, for instance MPs in a parliamentary debate, may make on the basis of (their understanding of) this speech and its context" (van Dijk 2005a: 66).

Though this formulation is quite general, van Dijk states that he intends to focus not on the semantic inferences a hearer may draw, but on the pragmatic (i.e. in his sense, contextually-driven) inferences the latter is able to derive from the linguistic stimulus, as he defines political implicatures as "implicatures that are specifically based on the political context" (van Dijk 2005a: 69).

We could first call into question what it means to characterise a context as political; by this, do we (vaguely) mean that the representations contained in that context belong the political sphere? Or that these may have a political impact on their addressees in that they can favour persuasion and lead to acceptance of certain standpoints? Van Dijk remains rather discrete on these issues, and we are left with little more than the intuitive yet vague idea that a political context is one where we find assumptions that have political significance in the world or that fulfil a political purpose.

More to the point related to implicature processing, it is striking that van Dijk does not resort to any of the available technical procedures to legitimate his use of the term 'implicature'. In Grice's account, implicatures are derived by observance or overt non-observance of the four conversational maxims (Quality, Quantity, Relation and Manner); if we go by Horn's model, we

need to explain implicature generation by calling in the Q- and R-Principles; if we adopt a Levinsonian approach, we will want to assess how the Q-, the I- and the M-Principles (respectively Quantity, Informativeness and Manner) are mobilised in the generation of implicit contents; if it is Relevance Theory we want to use, then we have to account for implicature derivation in terms of cognitive costs and benefits as regulating parameters of information-processing. But there is no account of how 'political implicatures' are technically derived in van Dijk's paper other than a mention of his theory of context models:

> "A more formal account would have to make explicit the precise context models of Aznar and the MPs, so that it can be shown how previous political knowledge, the representation of the situation and the mental model representing the semantic interpretation of this fragment all provide the information necessary to derive these plausible political inferences". (van Dijk 2005a: 73)

Even if we were to build and make explicit such a formal account, it is unlikely that we will get a thorough explanation of implicature derivation, because context models, as far as I know, do not provide any cause-to-effect modelling of how interpretations are arrived at, since they rather try to explain how knowledge is structured in our brains.[89] While such an approach is interesting in its own right, it does not explicitly address the link between the knowledge we have of the world and the meaning outputs we calculate on the basis of contextualised utterances. More precisely, context models do not supply an explanatory model of meaning construction, as they seem to focus on knowledge sets more than on the cognitive machinery involved in putting knowledge to use in dynamic mechanisms of information-processing in communication.

Van Dijk postulates that the notion of implicature is crucial in the analysis of political discourse:

> "political implicatures are what define also the political *functions* of the speech in the political process, and focus especially on Aznar's role as prime minister, party leader, as well as the legitimacy of his government and its international policies. At the same time, the implicatures have the function to derogate and attack the opposition in the public sphere. It is this *political* analysis of the speech that may be a contribution to the study of the political function of the speech in the political process." (van Dijk 2005a: 89)

Yet, it seems to me that van Dijk wants to endow the notion of implicature with a discourse analytical specificity that it cannot intrinsically bear. Pragmatists have studied implicature as a feature of natural information-processing which is independent of any particular use of

[89] See the notion of "K-devices", (van Dijk 2008a: 83), which van Dijk construes as "a central device that regulates the (non) expression of knowledge in discourse.

discourse; in other words, implicature denotes a standard feature of our communicative rationality.[90] The fact that implicature may play a role in political discourse analysis thus has nothing to do with implicature itself, but rather with the input implicature processing takes in particular cases, which is something van Dijk addresses with his theory of context models, but fails to make explicit in this paper. Consequently, I contend that the way contextual information is managed is more relevant to an account of political discourse than a description of implicature processing (which is incidentally not given in the article). Van Dijk fails to show this, as his paper emphasises the role of implicature instead of that of context. Implicature is traditionally approached as a formal procedure, and as such its study is relevant for political discourse analysis only to the extent that it enables the researcher to assess the particular ways in which context is fiddled with in order to achieve political aims. In short, implicature itself is not political, but the way it may be used can have political significance; this is not clear from van Dijk's exposition.

Having exposed some problems related to van Dijk's usage of the notion of implicature, I will now discuss his analysis of example (34) to show that it is also open to questioning. From (34), van Dijk argues that four implicatures, which I reproduce below in the second quote, can legitimately be drawn:

> "Thus, by making his own 'appearance' and 'report' for the MPs explicit in the very first sentence of his speech, he politically implicates that (i) he is doing his job as MP, (ii) he listens to the opposition and the country and hence is a good democrat, (iii) that the earlier critique against him is not or no longer relevant, and (iv) that those who formulated the critique – such as the oppositional PSOE – have no point." (van Dijk 2005a: 73)

I would like to discuss two issues related to the fact of terming (i) to (iv) implicatures. First of all, while I do not deny that propositions (i) to (iv) may be identified as relevant speculative interpretations of Aznar's words by the analyst – provided s/he engages in systematic and (perhaps overly) thorough contextualisation characterised by her/his own responsibility in providing all the contextual assumptions which are required to yield these propositions – there is no decisive evidence to assume that these interpretations correspond to the propositional contents the MPs who were listening to these words calculated and therefore there are grounds to question their status as standard implicatures (since implicatures are intended propositional contents). There seems to be a gap between the analyst's and the audience's interpretation, but

[90] See chapter 1 on Grice's rationality of conversation (§I.1.1.)

van Dijk gives no conclusive evidence to support the idea that these are equivalent, as he seems to straightforwardly assume.

Secondly, there isn't conclusive evidence either that Aznar, by formulating his utterance, intended to communicate all of these four so-called 'implicatures', as we do not see very well how the contextual assumptions required to derive (i) to (iv) are made mutually manifest by Aznar in example (34). While traditional neo- and post-Gricean accounts of implicature take into account, as van Dijk does, the idea that context can play a crucial role in implicature derivation (in that at least some assumptions which are required for proper processing of implicit contents are retrieved from the context of the utterance), another equally defining feature of implicature, which is absent from van Dijk's account, is its reliance on the relationship between meaning and intentionality.

Implicature was never meant to be taken as an analytical tool that would enable the analyst to list, by adding relevant contextual assumptions, several if not all possible interpretations of an utterance. Rather, it is meant to account straightforwardly for the way conversational participants manage and overcome semantic underdeterminacy by making assumptions about each other's intentions, beliefs, desires, i.e. each other's mental states.

Coming back to the example, I would argue that the four propositions van Dijk identifies can doubtfully be taken as four different or complementary *simultaneous* interpretations of what Aznar said and thus that they are not implicatures. Saying that (i) to (iv) are implicatures of the same input utterance would firstly entail that Aznar intended these four interpretations to come across at once, which is not plausible, since it is very unlikely that a single utterance will simultaneously mean four different things in the same speech situation, granted we stick to the classical account of implicature in pragmatic literature.[91] I contend that this is doubtful for obvious cognitive cost-related reasons and hence that (i) to (iv) are not implicatures proper, but weak implicatures (in Sperber and Wilson's sense) at best, even if this can still be debated, as we shall see shortly. What is certain is that they are possible interpretations requiring a great deal of contextualisation, and in this sense they could be relevant in the analysis. But the proper way to carry out this analysis would have to do without standard implicature, as (i) to (iv) do not seem to qualify as implicatures.

Let me consider the example more closely by giving an overview of the inferential work that would be involved in the derivation of (i) to (iv), in order to assess the implications of

[91] As van Dijk does not signal any theoretical departure from standard pragmatics in his paper, we are entitled to infer that he considers 'political implicatures' on a par with standard implicatures, and I contend this is why confusion may arise.

considering the latter to be implicatures. In a standard pragmatic account, the derivation of implicatures can pretty much be taken to follow some kind of argumentative structure, whereby they typically map an inferential path which builds on a combination of premises leading to a conclusion (the implicature *per se*). The derivation of the first of the four implicatures van Dijk identifies in (34) would roughly look like this:

(35) Aznar stands this afternoon before the Chamber in order to inform MPs of the Government's position (input utterance)

(36) People who stand before the Chamber and who inform the MPs of the Government's position are doing their job as MPs (contextual assumption)

(37) Aznar is doing his job as MP (= (i))

If (37) is indeed an implicature, then the contextual assumption in (36) is required to license the connexion we need to operate from (35) to (37). This would schematically be the way to analyse (34) as meaning (37) (or (i)). Now, this entails that the proper derivation of (37) is impossible without (36), which means that if van Dijk takes (34) to mean (37), he is *ipso facto* committed to assuming that (36) is present in the audience's cognitive environment (or context model) at the time the utterance is processed or that (36) is naturally and automatically brought about by the mere utterance of (34). While I would have no trouble agreeing with the possibility of the inferential path represented above, I still consider that evidence should be given in order to back up the assumption that (36) is part of the MPs cognitive environment at the time Aznar utters (34). Otherwise the conclusion that (37) is indeed an implicature is far from certain.

Perhaps this idea would gain more weight if we tried to spell out the more complex inferential steps involved in the derivation of (iv) (reproduced in (42) below):

(38) Aznar stands this afternoon before the Chamber in order to inform MPs of the Government's position (input utterance)

(39) The opposition has criticised Aznar for ignoring parliament and not making public the Government's position on the crisis with Iraq (contextual assumption)

(40) Standing before the Chamber in order to inform MPs of the Government's position amounts to making the latter public (contextual assumption)

(41) The opposition's critique no longer holds (implicit sub-conclusion)

(42) Those who formulated the critique have no point (= (iv))

If we want to ascertain that when Aznar formulated (38), MPs understood (42), then we need to accept that the latter must have had the assumptions (39), (40) and (41) readily available. But the grounds to do so are rather shaky; it is one thing to consider that the inferential chain taking

us from (38) to (42) is plausible and moreover valid, but quite another to assume that each of the intermediate steps was verified in the example van Dijk refers to, since one cannot but speculate about people's cognitive environments. In other words, the analysis should be limited to saying that (42) is a derivable implicature of (38) provided (39), (40) and (41) are fulfilled but also clearly stipulate that we have no way of ascertaining that such is the case. Of course, the researcher knows the context in which this discourse took place, and can speculate that Aznar's audience has in mind the fact that he has been accused by the opposition of not informing the parliament and the Spanish people about his foreign policy regarding Iraq. Yet, this remains speculative, and as a consequence so does the idea that (42) is a proper implicature. By naming it 'implicature', van Dijk is forced to accept that the inferential chain above has been verified, and I contend that this is far from evident. In van Dijk's terms, such an interpretation is likely to be reached only if the context models he attributes to the audience include all the assumptions which, combined to the utterance, would yield (42).

Perhaps an additional line of argument could help providing the discussion with a more plausible outlook, and it has to do with a difference Sperber & Wilson (1995) have established between implicatures and weak implicatures. Proper implicatures are traditionally taken to fall under the scope of full-fledged speaker meaning; however, alongside implicatures, there are additional implicit contents which can be derived by supplying additional contextual assumptions: those are also termed implicatures, albeit of the 'weak' variety (see Sperber & Wilson 1995: 197-202 and Blakemore 1992: 128-132). One of the main differences between implicatures and weak implicatures is that the first necessarily require that the hearer assumes that the speaker intended him to go through all the inferential steps involved in its calculation, while such is not necessarily the case for weak implicatures. To take Sperber & Wilson's example, a speaker implicates "No" when she answers "I would never drive an expensive car" to the question "Would you drive a Mercedes?", but could also be taken to weakly communicate something of the sort "I disapprove of any display of wealth", should the appropriate contextual assumptions be supplied. The latter interpretation can result from the hearer's mobilisation of further contextual assumptions, but does not require, in order to be successfully derived, that the hearer assumes that the speaker intended to make mutually manifest the particular contextual assumptions required to calculate said conclusion. In other words, the derivation of weak implicatures can be successful without resorting to speaker intention. This by no means entails that weak implicatures are completely independent of speaker meaning, for a speaker may very well *a posteriori* grant the validity of a weak implicature which is consistent with his beliefs. However, and this is the focal point of my re-interpretation of van Dijk's examples, just because a representation is consistent with one's beliefs does not mean that it was meant to be communicated. Accordingly, I do not take weak implicatures to be intended by the speaker to be

derived as the proper meaning of her utterance, and this is the main difference with standard implicatures. I would thus venture that propositions (i) to (iv) identified as implicatures by van Dijk in example (34) above are not implicatures, but at best weak implicatures, which are something completely different.

The classical definition of implicature necessarily requires that whatever content implicitly encoded be ostensibly meant by the speaker. Yet, in the example, even if (i) to (iv) can correspond to relevant interpretations, I would still proceed with a great deal of caution and refrain from attributing to Aznar the intention of conveying all four so-called 'implicatures' through the first sentence of his speech. Moreover, it would be very problematic to assume that a single utterance such as Aznar's has four different, and yet simultaneous, interpretations ((i) to (iv)), as van Dijk seems to advocate, simply because this would entail that addressees understood four different propositional meanings from one single utterance, which is rather implausible, given the costly cognitive work, time constraints and possible confusion it involves. Blakemore highlights a requisite of implicature which can chiefly summarise the objection to van Dijk's account of example (34): "A speaker who is aiming at optimal relevance must assume that the contextual assumptions required for the interpretation of the utterance are immediately accessible to the hearer" (Blakemore 1992: 128). I argue that with example (34), we have no guarantee that the contextual assumptions required to derive (i) to (iv) are indeed immediately accessible to the hearer. This is certainly an existing possibility, and this is what a cautious analysis of (34) should precisely stipulate instead of bluntly asserting that (i) to (iv) are implicatures. In that sense, van Dijk's analysis of example (34) is an interpretation more than an analysis of Aznar's words.[92]

Summing up this second objection, I would like to state that the point of van Dijk's analysis is not wrong *per se*, but that his use of the notion of implicature as a discourse-analytical tool is inadequate, as the contents he identifies as 'political implicatures' are more likely weak implicatures that the audience might make on the basis of their cognitive environments and the utterance, but without necessarily resorting to the assumption that this is exactly what Aznar meant by his utterance. After all, in order to infer from the first utterance in the example that Aznar thinks that the opposition's critique misses the point, one does not need to postulate that Aznar formulated his utterance as a way of getting his audience to recognise that he thinks that the opposition's critique misses the point (considering this would serve as a good enough reason to infer it). It consequently seems that it is not implicature we are dealing with here, but rather

[92] This echoes Widdowson's critique of CDA as interpretation more than analysis (see §II.1.1.2.2.2.).

weak implicatures which can be drawn independently of whether the speaker intends them to go through or not.

This last point naturally calls in the issue of speaker commitment. Recent research[93] has focused on whether implicatures do or do not involve speaker commitment, by trying to assess if the hearer's derivation of an implicature necessarily goes along with the speaker's endorsement of the assumption according to which she meant to communicate that content. The idea is that if the speaker denies the implicature, then it means that there was no implicature to be drawn in the first place, and that the hearer did not understand the speaker's intended meaning properly. Now, even if there is no absolute way to refute that Aznar could turn out, if asked about it, to endorse each of the four propositions (i) to (iv) van Dijk identifies, it is arguably very likely that he would deny having communicated them simultaneously *with that single utterance*.

To conclude on van Dijk's usage of implicature, let me insist on the idea that it is one thing to *a posteriori* list the *possible contextual implications* of an utterance, and quite another to justify that these were all implicatures both meant by the speaker and derived by the addressees. Pragmatic research is interested in this second option, but does not ambition to say much about the first, if only because, in the absence of contextual cues, an utterance may virtually give rise to limitless interpretations. In this sense, it could be said that van Dijk's usage of the notion of implicature does not do justice to the notion developed in Pragmatic research since Grice, as he seems to take for granted Aznar's intention and the audience's '*ad hoc*' cognitive environment.[94]

A final objection of principle can be opposed to van Dijk's account: it is impossible for an implicature to remain hidden and still have a particular effect, because i) if an implicature is derived, it is because it has been recognised as intended by the speaker (and thus its content as well as the fact that it was intended by the speaker is necessarily manifest to the hearer), and ii) if the implicature is not derived, then its content will not be accessed by the hearer, and postulating that a content a hearer did not derive at all can have an effect on him whatsoever would require significant, possibly empirical, input from research in psychology, which CDA does not provide. This was manifest in van Dijk's (2005a) example of Aznar's speech; the propositions he identified as political implicatures cannot be implicatures under the standard account. But what is more interesting is that an implicature, in abstract terms, even if it requires

[93] Cf. Morency, Oswald & Saussure (2008), Saussure & Oswald (2008, forth.).

[94] That is, only if we go by traditional pragmatic standards. I do not exclude the possibility of using a conceptual or analytical term in a sense which is different from its original meaning. But this requires some explicit mention of how and why the term is not used in its original acceptation; moreover a critique of the original account would be welcome in order to motivate a departure from it. Van Dijk does not provide this.

more or less costly processing effort to be derived (in that the hearer must mobilise contextual assumptions to retrieve the intended interpretation), can *never* be covert, thus it cannot escape the hearer's awareness; it is part of the definition of the implicature that it is necessarily meant to be communicated, and therefore *overt*. Therefore, any content communicated in this form, as long as the implicature is retrieved, will always be ostensively manifest to the hearer. Building on the implicit nature of implicatures in order to argue in favour of a *covert* transmission of information of ideological significance through implicature processing is incompatible with a standard pragmatic perspective. Deriving an implicature entails attributing an intention, and thus recognising that the speaker intended to convey that particular meaning; there can be no 'hidden implicatures', because by definition implicatures cannot be 'hidden'. A covert implicature is thus a contradiction in terms.

Even more to the point, to exclusively build on implicitness to characterise a manipulative attempt would be a mistake, since by definition manipulation is successful only if the hearer is not aware of the manipulative attempt (assuming that communicative implicitness is both ostensive and overt). Rather, it is covertness that is crucial to manipulation, not implicitness, and I think it is worthwhile pointing out that these are not the same thing. So while the content of an implicature may very well contribute to the success of manipulation,[95] it is counter-intuitive to assume that this is due to its implicit nature. Such a mistake, I have argued, is very likely to occur if we make use of the concept of implicature without first making sure that the way we use it is in accordance with its original account.

I have tried to show by commenting on his use of the notion of implicature and by recalling the original account of the term that van Dijk's own account of 'political implicature' can be seriously challenged and that he misuses the concept in his analysis. This, in my opinion, does not mean that the point of his analysis is far off. Rather, it indicates that the tool he uses is neither applied consistently nor calibrated to serve the purpose he wants it to serve, thus exemplifying Widdowson's claim that there is a tendency in CDA to overlook the detailed features of the analytical and conceptual tools it borrows from other disciplines.

In what follows I will evoke a contiguous critique that could be addressed to CDA on the grounds that the advocated multidisciplinarity may be driven by social concerns, namely a – perhaps too exaggerated – fascination for the non-literal, which can be detrimental to the proper use of such a heterogeneous analytical tool-kit.

[95] See example (69).

II.1.1.2.3.2. Passivisation as a typical CDA case

CDA's commitment to root out the delusion, to deconstruct the way ideologies and power structures operate through discourse most of the time in a hidden way, is reflected in its focus on the analysis of linguistic devices and mechanisms that allow the subtle communication of hidden, covert, or at least non-explicit, contents, with a tendency to use these terms interchangeably. The absence of certain elements (such as, for instance, the non-representation of syntactic and semantic agents in passive constructions), the presence of implicit contents (as in presuppositions and implicatures) are therefore typically focal points of the analysis. As a continuation of the preceding section, it is my purpose here to discuss briefly the nature of some other aspects of non-literal communication in order to assess whether their alleged covertness can reasonably be taken to play a role in the surreptitious propagation of ideologies and the (re)production of social asymmetrical relationships. I will defend the idea that it is not because a content is not explicitly formulated that it is necessarily deliberately hidden so as to escape the addressee's awareness and serve an ideological purpose, which is something CDA scholars tend to overlook, and that could be avoided should the original accounts associated to these notions be consistently applied.

Another example which illustrates a tendency to overemphasise the role of specific properties of linguistic expressions revolves around the notions of nominalisation and passivisation, as applied to the analysis of media texts in earlier works by the East Anglia school of Critical Linguistics. The canonical example of passivisation (cf. Trew 1979 and below) involves the representation of a clash between police and demonstrators, and it is claimed that a report of the event through an agentless passive construction – as the passive voice allows omitting the agent who is responsible for the action represented in the utterance – allows mystification on the grounds that the hearer would have to engage in additional cognitive calculations to work out agent reference. I contend that this, on the face of it, may be plausible, but only to the extent that we focus our attention on one single utterance taken in isolation.

There are indeed a number of reasons to believe that the story is far more subtle than it seems. First of all, utterances in media texts, including those containing passives, do not come alone. They are part of a more global narration, in which sometimes the representation of the event is accompanied by a photograph, and is most of the time inserted in a more elaborate narration or commentary, so as to fulfil editorial and reporting requirements of completeness and precision. It is thus quite straightforward to assume that readers do have at their disposal a fairly large array of sources of information to build up the context of interpretation for any utterance taken from a newspaper article. Let me take the classical example provided by Trew (1979) to illustrate my point:

(43) "RIOTING BLACKS SHOT DEAD BY POLICE AS ANC LEADERS MEET

> Eleven Africans *were shot dead* and 15 wounded when Rhodesian police opened fire on a rioting crowd of about 2,000"… (Trew 1979: 94)

Trew argues that in the opening sentence of the body of the article, as the passive voice deletes the agent, the latter is only weakly retrievable by implication from the temporal vicinity with the event represented in the following clause (the fact that police opened fire). While the point made that the agent is deleted cannot be questioned, the rest of the argument, namely that this deletion may deceive readers into incorporating a police-friendly ideology, and that this results from the non-processing of the agent of the action prompted by its literal absence in the sentence, can be challenged. First of all, the title of the article immediately preceding the sentence Trew analyses is formulated with a 'complete' passive form and specifically identifies the police as the agent of the killing, so I take it that readers would have to be extremely distracted not to make the connexion with the information contained in the immediate co-text, to the extent that it seems implausible to take "rioting blacks" and "eleven Africans" not to co-refer. Second, the way we conceptualise events denoting human action typically requires the representation of agents and patients, especially if what is at stake is the narration of a series of events. Whether these are mentioned or left out of the literal representation of the event in principle does not say much about the way the readers of that particular sentence will conceptualise them from a cognitive point of view; notably, it is not because something is not literally expressed that we will fail to process it, as the case of implicit information contained in implicatures or explicit (yet not literally expressed) information required for the fulfilment of truth conditions (such as the so-called *unarticulated constituents* identified by Perry (1986)) vividly shows. To take Perry's canonical example, when I say "It is raining", my addressee will automatically supply the non-literal deictics "here" and "now". So the fact that something is absent from the literal form does not necessarily entail that it is inaccessible, let alone remains unprocessed.

More to the point, would it be plausible to imagine a report of a riot, within the context of a news article, depicted exclusively as its result, i.e., as the representation that all of a sudden we find dead people lying around? That is, is it plausible to assume that readers, just because the agent is not explicitly mentioned, fail to represent the causal chain of events that led to this state of affairs and thus fall for the ideological trick (assuming there is one on behalf of the text producer)? It sure is a possibility, but it calls for empirical validation (for instance questionnaires), which Trew does not provide. In terms of an explanation of why the agentless passive should work the way Trew says it work, we are consequently left with the speculation

that readers are either very distracted or that their ability to conceptualise events is somewhat impaired.

A syntactic-semantic argument can be also called in here; the use of a multi-agent selecting verb semantically (and even formally) presupposes that argument positions have to be filled; just because they are omitted in the sentence does not mean that the addressee will not work out reference. I have mentioned above that in the context of a journal article, it is usually pretty clear and easy to figure out who does what to whom thanks to the context (including the co-text), so that supplying the adequate arguments should not be cognitively too costly. If readers could not work out proper argument assignment, then they would take the article to be obscure, and misunderstandings may occur. But this is certainly not the usual case. Consequently, from a syntactic/semantic point of view as well, the fact that the agent is omitted in passivisation is not sufficient evidence to maintain the necessary link between the absence of certain elements and the surreptitious communication of ideology. At best, one could conclude from this type of omission that the producer is ideologically biased; but this very different from assuming that addressees might fall for the trick and unwittingly adopt said ideology.

Third, as Trew notes it himself, it is possible to identify the agent by implication once the subsequent argument "Rhodesian police" is mentioned as opening fire on a rioting crowd; the identification of those who shot the "eleven Blacks", which are not mentioned through the agentless passive, seems to naturally follow from the representation provided in the second clause. Both the patient and the action are the same in both clauses and this could motivate addressees to straightforwardly assume that the absent agent in the first clause is the same as the explicit agent mentioned in the second. While there may be complex processes involved in the conceptualisation of events in terms of reference assignment, it seems that the claim Trew makes calls for some kind of validation, and that without it this remains questionable.[96]

Among other features of language critical linguists, and after them many CDA practitioners, traditionally studied, we find *nominalisation* – i.e., the fact of using a noun instead of the full description of an event where someone does something to someone else, thus reifying the event so as to, for instance, reduce the agent's responsibility –, *modality* – which allows the representation of events and situations with variable degrees of plausibility, as well as formulating wishes, orders, suggestions, etc. – and *lexical structure* – which draws on the categorisation of referents and discursive entities into specific sets, this categorisation being ideological itself. What all these devices have in common is that they represent the world in a

[96] See also O'Halloran (2003) for an analysis of the same example which draws on four different theories to show Trew's simplistic analysis.

particular way; they are consequently afforded the status of mental constructs. They also make these representations available through communication, allowing for the dissemination of ideologies. All communicated representations, as mentioned above, are socially constructed according to critical linguists, because they reflect a particular belief/value/ideological system; it is thus believed that once processed by an audience, such contents affect the latter's own beliefs/values/opinions. In particular, CL and CDA go by the idea that language is in this sense a means to shape other people's beliefs, values and opinions. To the extent that their exploitation in the analysis is similar to that of passivisation, they are susceptible of receiving critiques which are similar to the one I just exposed above on passivisation.

What I wanted to show with this discussion is that it may very well be the case that, taken in isolation, agentless constructions turn out to be misleading but nevertheless that these constructions seldom come in isolation – at least in media texts – and most of the time a context which allows disambiguation and referential saturation is readily available, either co-textually or from situational variables. So the fact that a linguistic item which is relevant in the representation of an event that took place in the real world is left out by a particular linguistic formulation does not necessarily mean that it will result in some kind of mystification, simply because other clues may make up for that omission (such as contextual co-reference). Assuming that passivisation as such leads to textual mystification is forgetting that readers are usually efficient information processors with an evolved ability to 'fill in the blanks' left by implicit formulations. As a consequence, the fact that something remains hidden in language is not a sufficiently clear or reliable indication of its misleading potential, which is something research in Critical Linguistics – and by percolation in CDA – tends to overlook.

II.1.1.2.3.3. Presupposition in CDA

A final linguistic phenomenon bearing non-literal status which I'd like to examine because it has been used in CDA research (e.g. Wodak 2007) is that of presupposition. I will only focus on one example which Wodak extensively comments on; it is taken from a speech delivered by then FPÖ leader Jorg Haider in 2001, during the election campaign in Vienna. As he talks about Ariel Muzicant, president of the Jewish Community, he utters a statement containing a dubious wordplay of anti-Semitic tone:

(44) "Mr. Muzicant: What I don't understand is how someone called Ariel can have so much dirty linen… I don't understand that at all, but I mean… he will certainly comment this tomorrow, won't he… but I am not frightened in these questions" (J. Haider on 28th February, Ash Wednesday Speech, Wodak (2007: 205), translated by R. Wodak).

In her analysis, Wodak claims that this utterance existentially presupposes that Muzicant made criminal moves, and goes on to elaborate on the fact that, according to Haider, such is the case because Muzicant "seemingly exploits the interests of the Holocaust survivors for himself and his business" (Wodak 2007: 216). She calls in the concept of presupposition to reveal the hidden features of this utterance that allow considering its racist nature.

Presuppositions are non-explicit contents contributing to the truth-conditions of a proposition that are usually identified by the negation test: when the clause containing the presupposition is negated, the latter is preserved; if that content is lost in the transformation, then it is not a presupposition. Let's take a closer look at the example, in light of this, by negating the principal clause:

(45) "What I do not do not understand is how someone called Ariel can have so much dirty linen"

which I will transform into

(46) "What I do understand is how someone called Ariel can have so much dirty linen".

Here the existential presupposition that Ariel exists is kept, and so is the existence of a "dirty linen" (which in the utterance is a metaphor for criminal activity). Moreover, it presupposes that, regardless of its amount, Ariel does have dirty linen. So in terms of the standard account of presupposition, the contents identified as such by Wodak indeed seem to qualify as presuppositions.

However, I would say that Wodak's analysis of the anti-Semitic implications of this example does not need to be laid out in terms of the role of the utterance's presuppositions: she says that from these presuppositions "a chain of anti-Semitic insinuations and associations are triggered […]: 'Jews are rich, are all businessmen, etc.'" (2007: 216). The anti-Semitic tone of the example does not lie in the presuppositions proper, but on the assumptions that can subsequently be derived from these presuppositions. And this does not depend on the presupposition alone: those assumptions require the presupposition to be combined with further background assumptions in order to be derived, which is a procedure which in this case takes presupposition as part of its input, but which is completely different from the presupposition itself. Thus, it is not by relying exclusively on the presence of presupposition that we are able to identify further anti-Semitic inferences. I would say that further background assumptions, which are not presuppositionally contained in the utterance – even though they are compatible with the utterance's presuppositions –, should be adduced to this end. In sum, it is as if Wodak wanted presupposition to explain more than what it can actually explain. In order to explain the anti-Semitic tone of example (45), therefore, presupposition should only be taken as part of the

explanation. Failure to acknowledge the limited (even if crucial) role of presupposition carries, I contend, the risk of compromising the soundness of the standard account of presupposition by wanting it to do more than what it was originally designed to do.

Looking more closely at example (45), it seems here that what is more significant than the presuppositional features of the utterance is its metaphorical potential. The conceptual metaphor associating the person, Ariel Muzicant, and the washing product has strong implications in terms of anti-Semitic significance once coupled with the information that purity and non-corruption (usually associated with washing products) are asserted as not being present. In that respect, the principal clause ("I do not understand…") highlights an inconsistency and implicitly calls for a judgmental statement such as "one would need to be very unintelligent not to see the inconsistency", which could contribute to triggering the acceptance of the anti-Semitic implication of the metaphor.

What I intended to illustrate with this last example is the fact that the primacy of the socio-political signification of the analysis may lead the researcher to use some conceptual tools too quickly and in a very narrow, sometimes erroneous, sense which i) departs from the standard account provided in the discipline it is borrowed from, and ii) makes the tool serve the purpose in a problematic way. I furthermore suggest that this would not happen should a discussion of the conceptual tool take place; if the researcher discusses the original account and argues in favour of a specification of the concept (either by loosening or narrowing it), then there is room to evaluate it and tentatively conclude on its adequacy.[97] But Wodak does not engage in such a discussion in the aforementioned paper, which could somehow appear as a misfortune, because we are left with the idea that she did not use presupposition consistently in her analysis, even if the point she makes is highly relevant from the perspective of its social significance.

More globally, I contend that the supremacy of the social purpose of the analysis, together with the seductive idea that non-literal components of meaning, which are inherently part of communication, may play a role in the (re)production of ideology, can yield problematic accounts which do not do justice to multidisciplinarity; in this sense, multidisciplinarity gone wrong may result in a loss of scientific credibility of the particular piece of research involved, to the extent that it could also affect the social intent of the analysis: after all, if the tools one uses are not consistently used, this may bear strong repercussions on the whole endeavour,

[97] See for instance Polyzou (in prep) for a discussion on presupposition as a useful tool in discourse analysis.

independently of the fact that the cause the analysis serves may be very well founded and legitimate from a moral point of view.[98]

To sum up this section, let me recall Widdowson's point that CDA practitioners may be "too anxious to make a political point, too quick to come to critical conclusions" (1998: 149-150): I suggest that the evident seductiveness of non-litteral components of meaning in communication may lead to such a conclusion, which in the end is detrimental to the scientific credibility of CDA's practice.

II.1.1.2.4. Discussion

From this discussion of CDA's practice as a multidisciplinary endeavour, it appears that it faces a wide range of criticism, particularly if we look at it from an epistemological perspective aimed at testing the validity of CDA as a scientific practice. However, I have also noted that the objections I formulated can easily be dodged if we consider CDA not as a theory but as a collection of approaches unified by a political and social agenda. While I do consider that the tension I evoked is problematic from the former perspective, I do not endorse the view which Widdowson defends, implicitly at least, namely that CDA wholly lacks any theoretical requisites and that CDA practitioners are somehow 'free-stylers' of linguistic analysis. I still see their point as relevant, even if the scientific grounds on which they operate remain questionable. A discussion of the above issues seemed nonetheless legitimate, for research in CDA usually does not regard these criteria to matter as much as the criteria which allow measuring its effectiveness as a socially-driven enterprise. In van Dijk's words, what matters most is CDA's "contribution to change" (van Dijk 1993: 253); in light of the points made above, it could hardly be otherwise, given the problematic implications of the tension between the scientificity and the political significance of the type of account it has accustomed us to.

The value of the approach is to be found on a separate level. CDA's agenda is avowedly social and political. Accordingly, it is foremost interested in data where ideological contents and discriminative relationships are relevant. As such, and beyond the analytical issues I expounded above, that is, beyond CDA's choices to explain what it is that makes discourse problematic in terms of the (re)production of ideologies and asymmetrical social relationships, remains a point of high social significance. But the fact that there are many risks stemming from epistemological

[98] This is actually also Sokal & Bricmont's rationale for criticizing the work of some continental postmodern thinkers; they explain in their (1997) book how these scholars fail to consistently use findings from other disciplines such as Physics and Quantum Mechanics when they apply them to the Social Sciences. In this respect, their endeavour is not merely gratuitously critique, but crucially points to the fact that such an inconsistent practice may lead to discrediting the whole account.

concerns does not *ipso facto* disqualify the researchers' intuitions about these issues. Hence, I assume that CDA can constitute a very valuable *heuristic* source of ideas in the analysis of manipulation and uncooperative communication; even if it is scientifically problematic in that the practice itself raises many questions of scientific legitimacy, CDA still provides a valuable collection of general insights on communication and an equally relevant collection of real data analysis for that matter.

While the question of a basic set of epistemological requirements is not relevant for the evaluation of CDA as a practice from the point of view of many scholars within the paradigm, there are nonetheless several common assumptions which all CDA practitioners do share. The following section tries to assess, always from a scientific perspective, the operability of such assumptions, which revolve around common conceptions about discourse, ideology and communication.

II.1.1.3. Concepts and claims in mainstream CDA

II.1.1.3.1. Language, meaning and ideology

As a means used to represent and communicate about states of affairs, language – and by extension discourse, taken as a particular practice involving the management of linguistic entities, either in production or reception – is considered by some CL practitioners to be necessarily biased, in that it automatically encodes the speaker's own representation of how the world is and generally prompts the hearer to accept it. This idea is contained in Trew's vision of any representation being subjected to the representer's own ideological bias, according to which "all perception involves theory or ideology and there are no 'raw', uninterpreted, theory-free facts." (Trew 1979: 95) – meaning no ideology-free utterances. As such, discourse is taken to always carry an ideological load. However, this conception echoes common sense only to a certain extent, and it remains furthermore questionable, if taken as an inherent feature of *all linguistic productions*. Fowler states that "it is the essence of representation that it is always representation from some ideological point of view" (1991: 85). This resonates with part of Habermas's claim according to which "language is also ideological" (*op. cit.*), but also, to a certain extent, Roland Barthes' provocative statement in his introductory lecture at the College de France in 1977:

> "*La langue, comme performance de tout langage, n'est ni réactionnaire ni progressiste; elle est tout simplement fasciste; car le fascisme, ce n'est pas d'empêcher de dire, c'est d'obliger à dire*".

> "Language, as the performance of speech, is neither reactionary nor progressive; it is simply fascist; for fascism is not to prevent from saying, but to compel to say". (my translation)

This conception shares with some of CLs claims the view that ideological positions are deeply embedded in language, and of course, meets significant consensus among scholars in contemporary social science as it highlights the power language is endowed with within modern, deliberative, societies. The problem attached to this perspective is that such a radical assumption leads to considering a *necessary* connexion between language and ideology.

Let me now state a number of reasons why the statement that language is inherently ideological should not be taken for granted without discussion.

A first argument against this view is that some rather commonsensical observations may jeopardise the claim. Imagine my Swiss friend Laszlo is checking the website of an online shop because he needs to buy a new monitor for his computer. Because he is not familiar with the Imperial Units system, he asks me "How big is a 15 inch monitor?", and I reply "I think there are about 2.5 cm in one inch". One would be hard-pressed to assume that my reply conforms to, let alone that it produces or reproduces, any ideology, simply because the manifest purpose of my utterance is to be helpful to Laszlo by being descriptively informative.[99] In other words, the goal of my contribution is to be helpful by providing information which is relevant to Laszlo's goal of getting a new monitor, and even if I'm representing a state of affairs of the world, which happens to be a construction of human beings (a measuring system), on the face of it the contents I'm conveying do not seem to be formulated from a particular ideological stance by default. In fact, there seems to be no essential difference between my utterance and the utterance "1 inch corresponds to 2.54 cm", found in a book or any document providing measurement equivalences, be it a book produced and published either within a Swiss, Chinese, or Argentinean community. What is more, I would also take it that in terms of informative content, such an utterance could very well have been generated automatically by a computer; this making no difference at all as far as ideology is concerned: the statement will remain informative and helpful provided someone uses it as such, and yet, it would be plainly wrong to assume that

[99] There is no obvious necessary ideological stance in providing measurements, or, say, in being helpful. Even if there were, this is totally independent of language (one can be helpful without communicating), and as such is irrelevant to the connexion between ideology and language. We could still consider that my reply is ideological in so far as I'm adopting a particular measurement system, consequently privileging it over an alternative system, possibly as effective as the one I chose. Still, the point is that my contribution conforms to Laszlo's expectations, given that I know that in Switzerland people use the metric system: in this example I am not discriminating, in the ideological sense, among measurement systems.

a computer is being ideological, for computers just have no intentionality nor mental states. The propositional content is in all of these cases the same, and whether there is ideology in it is irrelevant to the point of the exchange.

One may argue that this example is not illustrative of language's assumed inherent ideological features, since it represents a state of affairs of the world which does not involve agency nor patiency, and therefore that there is nothing in there to be represented from a specific ideological perspective. Let us then look at an example of potential ideological significance: imagine I witnessed a burglary in a shop, and that I'm reporting the events to a friend of mine. I say "As I was in the corner shop, a man hit the shop keeper with a baseball bat and ran away with the money". My statement is descriptive, and represents the event as I saw it happen before my eyes. Now, I could very well have used a passive construction, such as "the shop keeper was hit by a man with a baseball bat". Or I could also have opted for the agentless construction, "the shop keeper was hit", but in that case, my interlocutor would probably have asked me "by whom?", for the report would look incomplete in that case. On the moral side, I may very well strongly condemn the burglar, or not, for that matter, if we imagine I come from a society where burglary and violence are tolerated, or if I happen to be one of the burglar's relatives, and know that he is stealing as a last resort and that he hit the shop keeper to defend himself. The point is that all this has no *necessary* connexion with my report of the event. I may certainly choose to formulate an ideological statement by emphasising a particular attitude, but it seems always possible for me to report without *intending* ideology to get in the way, as it remains equally possible to embed ideological considerations in my report.

While there is no question about the idea that language *can* be ideological, it is clearly too radical and furthermore scientifically questionable to consider that it is so by default. Holding such a view would actually run contrary to the Hallidayan idea whereby people communicate for different purposes. In some interactions, the overarching and most relevant goal for communicating happens to be descriptive, thus rather objective. In others, this purpose may be ideologically-driven. The point I want to argue in favour of is that the mere existence of choice in structure and its corresponding meaningful counterparts cannot, in principle, constitute a sufficient reason to rule out the possibility of purely descriptive statements, because opting for a descriptive ideology-free statement is also one of those choices. Agreeing with Fowler, Trew and in the same vein with Barthes, would amount to denying the possibility that language might be used to express neutral descriptions (which is a possibility afforded by the Hallidayan system, on which CL and CDA heavily build on).

Another reason why the claim that language is inherently ideological should be questioned is epistemological. From a falsificationist perspective,[100] the scientific status of a claim depends on its falsifiable nature. If we can think of an experimental design which would demonstrate that the claim might be false, then it makes sense to consider that it is scientific: we can assess it and test it to determine whether it survives experimentation or not, and whether it should be kept in the account or not. A claim such as "language is inherently ideological" is problematic on at least two counts. First, it establishes a necessary connexion between language and ideology, which poses the problem of triviality: if, by definition, all utterances are ideological, there are grounds to question the relevance, for a linguistic account, of addressing the relationship between language and ideology; moreover, if language is inherently ideological, then all possible utterances are ideological, which makes ideology an under-informative property: the only difference driven by ideological motivations we could find between stimuli would be qualitative; this has little to do with linguistic structure as it relates to the belief systems the utterances refer to, notably in terms of the identification of such and such ideology. This is independent of language itself: rather, we could say that language points to ideologies, which may anyway exist independently of language. Second, one would be hard-pressed to provide an experimental design that would falsify the claim; if by definition, language is ideological, no instance of language devoid of ideological features is imaginable. What would a non-ideological utterance look like? This possibility is ruled out by the claim itself, therefore revealing that it could not have been derived from scientific observation, and that it only results from a theoretical postulate. As a consequence, the claim "language is inherently ideological" cannot afford any prediction of a single case that would empirically invalidate its adequacy as a scientific statement.[101]

We can certainly assume that ideology is effectively disseminated through communication. But it is quite different, and in my view scientifically unfruitful – according to falsificationist scientific standards at least – to assume that all linguistic productions are ideological and to take this for granted.

A last argument can challenge the view according to which language is inherently ideological, and it can turn out to be highly problematic for CDA practitioners should they radically adhere to it. The argument can be put on a par with the paradox of epistemic relativism, which stems from an assertion of the type "the truth of any statement is relative to the values of the epistemic community its speaker adheres to". If such a sentence is true, the universal quantifier ('any')

[100] See Popper (2009 [1963]).

[101] I am grateful to Louis de Saussure (personal communication) for providing the bases of this discussion.

captures all possible utterances and thus 'universally' establishes the relativity of their truth as it were; however, the statement also means that the truth of that same statement itself is relative, thus yielding the paradox that the statement is universal by its form but particular and relative should we take its content seriously, which is self-defeating: if true, the statement would turn out to be false. When one asserts that language is inherently ideological, one must be prepared to accept that all statements are ideological, including this one. If statements are ideological and if ideologies are relative to epistemic communities, then so are statements, and someone asserting that language is ideological would end up facing the same kind of paradox as that of epistemic relativism.

But beyond the internal contradiction of this *aporia*, it bears serious implications for the practice of CDA as a scientific endeavour, i.e., as a rigorous discipline destined to produce reliable knowledge claims. CDA's purpose is to analyse how ideology, power and social relationships are produced and reproduced in discourse; the way it goes about in order to reach this aim is through linguistic analysis, which itself is a particular type of discourse consisting of sets of statements. As statements are ideological, it follows that CDA is itself ideological. Thus, under these standards, any analytical result CDA might reach will be relative to a certain epistemic community, and therefore partial. As Widdowson puts it,

> "discourse is a set of social values which we have to subscribe to as a condition of using language at all, so when we use language we do so as representatives of some socially determined discourse or other. Critical discourse analysts cannot of course claim any exemption from this general condition: they too are socially constructed to see things the way they do, and this, on their own account, renders them incapable of conducting any detached analysis at tall. All they can do is interpret other discourses on their own terms. Thus, paradoxically, they can only pursue their programme of ideological exposure by denying their own fundamental beliefs." (Widdowson 1998: 148)

It follows from this that either the claim according to which language is inherently ideological is false, or that CDA is, as Widdowson believes, interpretation but by no means analysis, which compromises its scientific credit.

In light of the above arguments, I am led to conclude that language is not inherently ideological. Nevertheless, this claim has no consequences as to the possibility of producing and interpreting ideologically-loaded contents. Ideology through language is possible, yet not necessary. This is a distinction that CDA has not made so far, and in my opinion can also be taken as a consequence of the primacy of social significance over epistemological matters.

II.1.1.3.2. Further elaborations

If language is not straightforwardly ideological but can nonetheless be a means of production and reproduction of ideology, we need to assess the way ideology can be accounted for, should it be accountable for at all within a broad theory of communication. In particular, we need to define ideology and locate where it is to be found within communication, which calls for an examination of the relationship between ideology and meaning.

Van Dijk defines ideology as "the foundation of the social representations shared by a social group" (van Dijk 2005b: 3). Leaving aside the vagueness of the term 'foundation' (what exactly is a foundation of social representations? does it bear a causative status over representational processes? who founds them? how? is a representation social in its scope, or in its nature?), it appears that ideology is of a representational nature. Having said this, we now need to evaluate whether it bears communicative status, that is, whether it is meant to be manifestly communicated in ideologically-loaded discourse. That is, when a speaker produces an ideologically-loaded utterance, does she intend her hearer to recognise that such is the case? Another way to ask the question would be to inquire about whether ideologies in discourse can take the shape of components of meaning (i.e., explicatures, implicatures, presuppositions, weak implicatures, etc.) and thus be accessed by addressees as propositional contents. If such is the case, then ideology would in principle be accountable for within any pragmatic theory, as it would constitute a particular aspect of meaning.

But there is a glitch. Assuming that a text producer is ideologically driven and indeed wishes to utter something of ideological relevance, is it in her/his interest to publicly endorse that she/he is being ideologically driven? It depends on whether said ideology is accepted or not by the audience. In cases where the speaker's ideology is consistent with the audience's, whether the speaker's ideology is manifest or not is not really an issue (except if one is interested in the social effects of such a communicative setting, such as for instance increasing complicity between conversational participants). In cases where the audience does not adhere to the values contained in that ideology, and if the speaker intends her ideology to have an effect on the audience whatsoever, then the ideological tone of her words must remain covert, which is a situation resembling that of manipulation, deception and the like (see chapter 1). Under these conditions, ideology does not bear a properly communicative status, as it is not meant to be recognised by the audience as intended by the speaker. So when ideology is hidden – which are typically the cases which CDA is most interested in – one can only speculate about the speaker's concealed intentions.

CDA goes by the assumption that it is through the study of text and talk that analysts can hope to gain more or less direct access to ideology, thus envisaging the possibility that ideologies can

take a propositional form through communication – granted indeed that there is a 'linguistic' way to go about them. Let us first inquire about the way ideologies are defined. According to van Dijk, it appears that ideology is distinct from language, even if some would probably venture that ideologies provide the structure of language. What can safely be argued (see van Dijk 1995: 244) is that ideologies bear cognitive status since they refer to sets of beliefs and values accessed on a representational level – this does not threaten in any way the validity of the claim according to which language may be used to express ideologies in a number of cases. Interestingly, and contrary to what most CDA scholars defend, there is room to argue that if ideologies are sets of representations, knowing that they may be socially shared has by definition no incidence on i) their nature as sets of beliefs, and ii) the interpretive mechanisms by which we process an utterance in which ideology is embedded (the variable here is thus to be found in the content, not in interpretative mechanisms proper). Therefore, a question which should occupy researchers in discourse analysis, more than the issue of finding linguistic means of identifying ideology, is rather to establish whether there are good grounds or not to consider that the speaker's utterance is ideologically loaded, particularly by assessing whether the purported ideological load is relevant to the interpretation of the utterance. The fact that these happen to be socially shared sets of beliefs is not crucial in the comprehension process. As far as the account of meaning is concerned, it can thus only be *descriptively* (as opposed to explanatorily) relevant to consider the social aspect of ideology.[102] Put somewhat differently, the issue is therefore about finding ways justifying that someone exposed to the verbal stimulus will actually work out its ideological meaning; it is thus about providing reasons supporting the particular interpretation the CDA practitioner associates to the verbal stimulus.

What I am advocating here is a shift to the hearer's perspective; of course, the analyst can come up with many interpretations of a given utterance, and some of them will turn out to be ideologically-loaded while some others not, depending on the context. But, as Widdowson puts it, the job of the analyst is not to come up with an interpretation and to try to find confirmation for it in the text; rather, it is to come up with potential interpretations and to specify under which contextual conditions the latter are licensed. For this, we need a proper theory of meaning that is able to accommodate some notion of speaker commitment.

(Micro-) linguistic analysis is concerned with the representations a given utterance may convey, and within the field of Cognitive Pragmatics, focus is explicitly set on what an audience may be

[102] Of course, the fact that a speaker adheres to an ideology influences the contents she encodes. However, it is only at the hearer's end of the communicative process that it makes sense to wonder about the relevance of the ideological load of an utterance, since an ideological statement in the social vaccum just has no ideological significance if no one is there to interpret it.

entitled to infer from a speaker's production. As a consequence, there is only a restricted number of propositions that can, *in a given context*, be inferred from an utterance. Recent work on the notion of speaker commitment with respect to the semantics/pragmatics interface (Morency, Oswald & Saussure 2008, Saussure & Oswald 2008) considers that hearers attribute commitment to speakers as they process the latter's utterance meaning. The main purpose of such an account is to provide a model of how hearers – online, as they process information – establish which explicit and/or implicit contents a speaker can be said to have endorsed, or taken for granted, in making her utterance.

Now, the thorny but crucial question we end up facing again when we address the issue of ideology from such a perspective is the following: is ideology a component of meaning, i.e., part of an explicature and/or an implicature communicated by an utterance, or, if not, what is it? Inasmuch as ideology is about beliefs and conceptual contents that are subject to representation, it seems plausible to assume its connexion with meaning: considering that ideology is an exclusively structural feature of language would be very hard to defend, since one definitional aspect of ideology is its referring potential.

Unless the point of an utterance is to provide information, somehow metalinguistically, of what an ideology is, I hypothesise that ideologies are neither explicit nor implicit communicative contents. First of all, they are believed to be sets of representations (beliefs, values, norms, etc.), and both explicit and implicit contents are limited: it would be counter-intuitive that a bank slogan is conveying the whole capitalist ideology (even if it is probably the case that traces of it can be read in the message). Yet, a component of meaning, whether explicit or implicit, cannot be a whole system of beliefs/values (and consequently cannot be an ideology).

Second, there is a difference between some aspects of meaning which the hearer can safely attribute to the speaker, and other contents which can be quite independent of the speaker's intended meaning, but which may nevertheless be relevant. In order to illustrate this, let us take a series of representations that one could possibly infer from the following exchange:

(47) a. Lucinda: "Do you want some coffee?"

b. Laszlo: "I don't drink coffee because coffee producers exploit workers".

(48) Laszlo does not want coffee.

(49) Laszlo thinks Lucinda should not buy coffee.

(50) Laszlo disapproves of coffee producers.

(51) Laszlo is left wing.

(52) Laszlo wants to make me aware of social inequalities

While (48) can straightforwardly be taken to correspond to what Laszlo meant by uttering (47b) and (49) represents an additional interpretation of his utterance, provided contextual clues allow for it, whether (50) or (51) can be considered to be meant by Laszlo has no real incidence on the derivation of the former. In fact, neither (50) (51) nor (52) require to be recognised as intended by Laszlo to be derived by the hearer, unlike (48) and (49), which are arguably precisely about the contents Laszlo *intends to communicate* to Lucinda. (50), (51) and (52) are assumptions one forms on the basis of a contextualised linguistic stimulus, but they are not necessary for the making of meaning. They are assessments of what Laszlo's utterance may tell about those beliefs and complex thoughts which he entertains independently of the issue at stake and which are not necessary for comprehension to take place, even if they could be conceived, following Saussure (2007), as higher-order inferences. What is more, as they require the mobilisation of specific contextual assumptions which Lucinda can only, with a variable degree of speculation, be taken to make manifest, they most likely are weak implicatures.[103] As far as ideology is concerned, if derived, (50), (51) and (52) may lead Lucinda to consider that Laszlo's utterance conforms to a certain ideology, but these conclusions are accidental, because we cannot say that they have been intended by Laszlo in the exchange we are considering, even if, all other things being equal, we could infer them independently of the pragmatic and functional content of Laszlo's utterance. In fact, Lucinda may or may not see that Laszlo's reply is ideological; this has no incidence on her understanding of his refusal to drink coffee.

This basic illustration suggests that processes of meaning construction do not straightforwardly encapsulate ideological aspects of utterances. The fact that an ideological representation which is independent of the actual meaning of the utterance may be derived by a hearer suggests that meaning and ideology are two separate things that do not necessarily entail one another. One way to characterise this difference would be to say that representations of meaning (of the explicature/implicature sort) are necessarily bound to speaker commitment, while ideological contents are not. This is compatible with the idea that we may very well take a statement to be ideological even in cases where the speaker does not intend to deliberately convey an ideology (it just happens that we read ideology into it). Ideological statements may turn out to be deliberate or not, but the way we assess this is different from the way we can safely assess whether a speaker has intended a content to be derived by the hearer, because we can always say that a statement conforms to an ideological system, even in the absence of any speaker. In other words, envisaging ideological communication from the angle of intentionality is probably not the best way to go, to the extent that ideology is a matter of reference which can operate

[103] See §II.1.1.2.3.1. on van Dijk for a discussion of the notion of weak implicature.

without intentionality entering the picture: it is the particular relationship between a content and a system of beliefs it relates to that makes the former ideological. Without such a system of reference, it would make no sense to speak of ideology. In turn, this means that a pragmatic account should not be foremost concerned with ideology, since ideological representations are not necessarily intended by speakers (in fact, it can also happen that speakers are completely unaware of their own ideological bias)

Ideological contents, contrary to explicatures and implicatures, are not the point of a given communicative exchange – except in metalinguistic communication specifically focused on ideology. Ideologies are sets of assumptions that hearers derive about the speaker's mental states and beliefs, which are contingent to the actual meaning of her utterance, whatever that may have been. In ordinary conversation, it seems that ideology *comes along* with meaning but is not the object of meaning itself. This suggests that ideologies are part of the context of utterance interpretation, but that they are not specifically the focus of utterance meaning.

Studying ideology from the linguistic point of view supposes that language contains its traces, which is an idea that we can easily defend to the extent that natural languages provide us with ways of expressing our subjectivity, our personal values and beliefs. This being so, it does however not automatically allow us to study ideology in language practices. First of all, I have noted that ideology is probably not accountable in terms of meaning, because unlike speaker meaning, ideology is usually not the point of an utterance. Additionally, there is a strong argument compromising altogether the study of ideology from a pragmatic perspective, and it has to do with the fact that ideology seems hard to situate on either sides of the semantic/pragmatic interface. Here is how it goes.

Assuming that ideology is something that is communicated supposes that it is either related to the linguistic code or to the speaker's intention. If ideology were linked to the linguistic code, this would make it a semantic issue, which is not plausible given that the same utterance may be ideologically loaded in some contexts but not in others. Moreover, this would make the account rely on a code-model, the inherent limits of which have been clearly shown by Sperber & Wilson (1995); additionally, it would be highly reminiscent of the far too radical Sapir-Whorf original hypothesis. If ideology, on the other hand, were linked to the speaker's intention, we would be extremely hard-pressed to account for it with available pragmatic tools to the extent that ideology is not a component of meaning standard pragmatic theories may be able to account for. It is not a simple propositional content such as an implicature or an explicature, and as such it usually cannot be speaker meaning.

Identification of a given ideology in a given utterance is therefore a very difficult task, because it is highly contextual, and because the speaker can always deny having produced such a statement

to the extent that it was only peripheral to the meaning she intended to communicate. The most cautious of all positions would thus be to consider that ideology is something lying in the eye of the beholder: if we qualify a text as ideological, then we should be prepared to defend our claim with strong arguments, in order to prove that the speaker can actually be taken to be committed to the purported ideological load of her utterance. If such arguments are shaky, which happens more often than not, then we should adopt a more charitable stance, by acknowledging that we may have read too much into the utterance.

The real strength of an account of ideology would be to be able to predict which statements will be ideologically loaded and to be able to show that a given statement is ideological. As I have tried to show, this is unlikely for reasons related to the very nature of ideology: it does not bear communicative status and it is not associated to particular linguistic forms. Therefore, we cannot definitely decide on the reasons of our ideological interpretation of a given message; furthermore, we can justify the intentionally ideological tone of an utterance only with great difficulty. To take the example of 'ideological' passive constructions, perhaps the speaker using them is trying to meet stylistic constraints that made her eliminate certain syntactic and semantic agents and privilege the passive form; perhaps she particularly likes that linguistic form; perhaps she used a passive form only by chance; perhaps, even, the utterance is ideologically loaded without the speaker even being aware of it. In this last case, would we say that the speaker is being ideological? It would presuppose too weak a notion of ideology: being ideological is linked to being dogmatic; the ideas forming the grounds of an ideological or dogmatic statement are known as non-questionable representations.[104]

My discussion of ideology, though only peripherally related to the topic of manipulative communication, is aimed at showing that CDA faces great challenges in terms of explanatory power when it comes to dealing with ideology. Notably, I contend that an account destined to identify ideology in linguistic forms is not scientifically powerful enough as long as we do not have a theory allowing us to clearly identify the conditions under which ideological contents can be safely attributed to speakers. Pragmatic theories of meaning cannot be of much help to the extent that ideology is usually not meant to be recognised as deliberately being embedded in the content of verbal utterances, which in my view constitutes one of the major drawbacks of a pragmatic assessment of ideological contents. For these reasons, I will not focus on ideology in the present work and shift my attention to other issues relating to manipulative communication.

[104] Saussure, personal communication.

II.1.1.3.3. Discourse as a social practice: issues of causality

Most if not all CDA practitioners agree on the definition of language as a social practice, and consider that, as such, its study cannot do away with the contextual complexities involved in its production and reception: "Describing discourse as social practice implies a dialectical relationship between a particular discursive event and the situation(s), institution(s) and social structures(s), which frame it" (Fairclough & Wodak 1997: 258).

CDA's claim is that in order to obtain a comprehensive and adequately informed picture of communication between human beings, it is necessary to consider the global context (spatial, temporal, political, social, cultural, etc.) in which it takes place. Language cannot be abstracted from its external influences, in so far as people engage in communication with their own goals, desires, interests and so forth – all these being embedded in more global social structures, which have an influence on interaction. If standard 'non-ideological' or 'non-strategic' communication remains nonetheless possible (as I have ventured in the preceding section), CDA will privilege the study of cases in which these standards are not observed, focusing on asymmetrical configurations of interaction leading to social inequality. This stance departs from the study of language as a competence, as Chomsky would put it, and focuses on the broader picture, with the inevitable consequence that some aspects of language (e.g. the cognitive aspects involved in its processing) may be neglected (see Chilton 2005a and a discussion in the paragraphs immediately below).

The definition of language as a social practice is based on two main ideas. First, on the now widely accepted pragmatic view that in interaction, people *act* in addition to exchanging contents. This conception traces back to the philosophical foundations of Speech act theory, in particular to the rejection of the 'descriptive illusion' (Austin 1962), and stipulates that discourse *is* action, what is more, *social* action, whereby uttering a content is already performing an act which affects and is affected at the same time by the social conditions in which the interaction takes place. The second idea behind the conception of language as a social practice is thus the view that communication is embedded in higher-level structures of interaction, which are subjected to socio-cultural (so to say, *external*) constraints. The value and interpretation of an utterance is defined from the identification of the particular type of action it performs, itself subjected to social constraints. Under this view, external influences on language determine its production and reception. CDA considers that the interpretation of any speech acts inevitably draws on its global context: an order will not be equivalent if shouted by a lieutenant to the troops or uttered by a parent to a child, because the circumstances in which each of these are performed are significantly different, and in turn the social implications of these acts might

differ greatly. Accordingly, the analysis should make explicit the social conditions under which any speech act is performed.[105]

In a (2005) paper, Chilton raises some serious issues related to the conception of language as a social practice. In particular, he discusses the three main tenets of CDA (adapted from Chilton 2005a: 22):

- discourse is social action (or 'social practice')
- social action constructs social reality (objects, situations, identities, social relations…)
- discourse is the use of language

If social action is discourse, and if social action constructs reality, then it follows that discourse (and *a fortiori* language use) constructs social reality. In addition to echoing the standard Sapir-Whorf hypothesis, this conclusion entails that discourse stands in some kind of causal relationship with social reality, insofar as the notion of 'construction' semantically entails that some agent does something to some patient in a causal fashion, for without the intervention of the agent, the state of affairs referred to in object position could not be asserted as the result of the state of affairs referred to in subject position. Now, there seems to be an interesting similarity between the terms 'social action' and 'social reality'; more precisely, the set of that which is 'social action' could very well be considered to be a subset of that which is 'social reality'; CDA provides no obvious reasons to consider that this should not be the case. Yet, if this holds, we would probably end up with a tautological statement such as 'social reality constructs social reality', or, since discourse is social action, which itself is part of social reality, 'discourse constructs discourse', and, drawing on Chilton's third statement above, 'language use constructs language use'. All these statements can to some extent be tenable – and Chilton notes that some postmodern frameworks are indeed quite comfortable with the idea that everything *is* discourse –, but they are so only at the expense of scientific informativeness. Saying that discourse constructs discourse does not inform us about the functioning of processes by which discourse produces whatever it is that it produces. What is more preoccupying, according to Chilton, is that these basic tenets lead to the conclusion that discourse ends up "causing itself" (Chilton 2005a: 23), in which case, again, informativeness is lost, this in turn being highly problematic for the explanatory power of the account.

Following Chilton's arguments, we end up with a quite dissatisfying picture of the scientific value of CDA: together with the arguments I exposed above regarding the tension between

[105] This corresponds to the explanation stage in CDA.

scientific and social concerns (§II.1.1.2.2.), it seems that beyond analysis, CDA has not made the effort to legitimise its practice in epistemological terms.

Chilton's conclusion derived from the three basic tenets of CDA formulated above holds only if we either assume a causal link between society and language, or if we do not make a distinction between two levels of discourse, namely, in O'Halloran's terms (2003) between *discourse*$_1$, the discursive practice of a text consumer, i.e., the interpretative stage of discourse processing by language users, and *discourse*$_2$, namely the socio-cultural practice, which refers to the more global system in which discourse$_1$ is embedded, and which is able to influence the latter. Let me elaborate on these two conditions.

II.1.1.3.3.1. Causality in CDA

The first of these conditions, namely that of considering a *strong* causal relationship between language and society, would never be endorsed by CDA practitioners, because it is not part of their research programme to allow such relationships to be the focus of the account. Wodak[106] explains the weaker status of causality in the social sciences by the impossibility of using the same standards that hard sciences use, notably in the empirical field: she notes that an experiment under standardised conditions (e.g. in Physics) is radically different from any experiment on genuinely spontaneous human behaviour, because the latter is never repeatable in precisely the same way, given that conditions cannot be controlled as they are in an experiment involving inanimate objects, for instance. A superficial review of some mainstream CDA literature on the relationship between language, mind and society reveals that the lexicon of causality is much 'mitigated'. Below are a few quotes illustrating this claim:

> "Describing discourse as social practice implies a dialectical relationship between a particular discursive event and the situation(s), institution(s) and social structures(s), which frame it. The discursive event **is shaped** by them, but it also **shapes** them. That is, discourse is socially **constitutive** as well as socially **conditioned** – it **constitutes** situations, objects of knowledge, and the social identities of and relationships between people and groups of people. It is constitutive both in the sense that it **helps to sustain** and **reproduce** the social status quo, and in the sense that it c**ontributes to transforming** it. Since discourse is so socially **consequential**, it gives important issues of power. Discursive practices may have major ideological **effects** – that is, they **can help produce** and **reproduce** unequal power relations between (for instance) social classes, women and

[106] Personal communication.

> men, and ethnic/cultural majorities and minorities through the ways in which they represent things and position people." (Fairclough & Wodak 1997: 258, my bold)
>
> "More specifically, critical discourse analysts want to know what structures, strategies or other properties of text, talk, verbal interaction or communicative events **play a role** in these modes of reproduction. [...] Hence, special conditions must be satisfied for such discourse properties to **contribute to** the reproduction of dominance." (van Dijk 1993: 250, my bold)
>
> "Although embodied in the minds of individuals, social cognitions are social because they are shared and presupposed by group members, monitor social action and interaction, and because they **underlie** the social and cultural organization of society as a whole." (van Dijk 1993: 257, my bold)

The terms used by Fairclough & Wodak and van Dijk in characterising the relationship between language, cognition and society appear to be quite vague, if understood in a causative perspective: when one says that something is constitutive of something else, does that imply a strong causal relationship? Similarly, what about asserting that something underlies something else, or that something contributes to transforming something else? CDA practitioners do not engage in a systematic reflexion upon causality as such in the social sciences and their work is reluctant to identify causal relationships because they do not believe that these can be asserted, given the constantly changing and heterogeneous nature of human behaviour. Hence the lack of commitment to causality, understood as a deterministic factor.

However, it seems relevant to notice that expressions such as 'underlies', 'constitutes', 'plays a role' 'shapes', helps to produce/sustain', 'contributes to transforming' do capture a semantic load of causality. If one asserts that a state of affairs A 'underlies', 'constitutes', 'shapes', 'helps to produce/sustain', 'contributes to transforming' a state of affairs B, one must be prepared to accept i) that the relationship at stake is relevant to *explain* why the state of affairs B came to be given the presence of the state of affairs A, and ii) that the state of affairs B would probably be inexistent or at least different if the state of affairs A had not occurred (for if such were not the case, discussing state of affairs A with respect to state of affairs B would be uninformative and possibly irrelevant). Furthermore, any discussion about (even weak or underspecified) causal relationships between states of affairs in the world supposes that we consider that things do not happen by mere coincidence; the rationale of any scientific account, as a matter of fact, lies in the explanation of phenomena, that is, in the identification of causes and effects. So even if CDA practitioners do not acknowledge an explicit commitment to causality between language, cognition and society, it still seems that they do express causal relationships, albeit in an obscure and, according to Chilton (2005a), circular sense.

II.1.1.3.3.2. Two notions of discourse

The second condition under which the statement 'discourse ends up causing itself' is indeed problematic relates to the complexity of the notion of discourse. While it seems obvious that any statement stipulating that an object its own cause seriously challenges our logical intuitions, CDA deals with the difficulty, though seldom explicitly, by relying on a complex and multi-faceted notion of discourse.[107] In order to avoid the logical inconsistency Chilton expounded, CDA adopts different definitions of discourse so that the difficulty implied by a statement such as 'discourse ends up causing itself' can be dodged. Here are a few insights on what discourse is taken to refer to in CDA.

Wodak & Reisigl (2009: 89) consider that discourse can be identified by relying, at least in part, on a thematic criterion: "We consider discourse to be […] related to a macro-topic […]. Thus, we regard […] macro-topic-relatedness as [a] constitutive [element] of a discourse". This is consonant with Kress' own take on discourse, according to which "a discourse provides a set of possible statements *about a given area*, and organises and gives structure to the manner in which a *particular topic*, object, process is to be talked about" (Kress 1985: 6-7, my italics). So, according to these first hints, discourse can be considered as an object which is somehow homogeneous in its topic. Under this view, discourse is a meaningful entity whose boundaries are defined in terms of the topic it deals with. We can therefore distinguish discourses on the basis of the different issues they address: the discourse of global warming, the discourse of civil rights, the discourse of abortion, and so on.

As far as the nature of discourse is concerned, some researchers define it as a consistent set of assumptions (beliefs, values, etc.) about the world. This corresponds to O'Halloran's (2003) definition of discourse$_2$ as bound to a specific socio-cultural context, and echoes Foucault's own definition of discourses as "systematically organised sets of statements that give expression to the meanings and values of an institution. For him, discourses define and delimit what is possible to say and not possible to say" (O'Halloran 2003:11-12). This is also Kress' take on discourse: he contends that "a discourse provides a set of possible statements about a given area, and organises and gives structure to the manner in which a particular topic, object, process is to be talked about" (Kress 1985: 6).

Discourse is also defined in CDA as the coherent understanding the readers makes from a text, which can be put on a par with the practice of understanding. Discourse, in these terms, is a process. This is what O'Halloran refers to as discourse$_1$. However, such a practice can be guided by some kind of 'higher-order' discourse, i.e., by the set of values and beliefs that will provide the

[107] Many thanks to Majid KhosraviNik (personal communication) for pointing this out.

perspective from which an individual will approach a given text – in other words, what O'Halloran calls discourse$_2$.

With his socio-cognitive approach, van Dijk defines discourse as knowledge sets stored in memory that will determine the production and interpretation of linguistic stimuli. In this sense, the term discourse denotes some kind of cognitive phenomenon: "language use (discourse) is, as the tenets of CDA assert, connected to the 'construction' of knowledge about social objects, identities, processes, etc." (Chilton 2005a: 23), and this is something that can only happen in people's minds.

Taking into account the many meanings that the term discourse can have, the problem of circularity unveiled by Chilton (2005a) ceases to be a problem, to the extent that the circularity of the argument is defeated: the argument is no longer circular because the term discourse refers to different things in its different occurrences. However, under these conditions another problem arises: with so many different acceptations of the term discourse, as noted by Widdowson (1995), terminological issues compromise the scientific instrumentality of the notion.

It seems therefore that one way of avoiding Chilton's charge on circularity would be to dispose of a clear and limited set of definitions of discourse. But CDA does not provide the type of terminological clarity one would hope to find. As a consequence, I will not retain CDA's framework for my analysis: its multi-faceted conception of discourse is not in my view suited to the type of account I am looking for.

II.1.1.4. Conclusion

In this section I have discussed several aspects of CDA, both in theoretical and practical terms. On the practical side, I have tried to show that the social commitment of CDA can be detrimental to the scientific thoroughness of the account: the hierarchically superior goal of making a point of socio-political importance can sometimes compromise an adequate application of analytical tools borrowed from neighbouring disciplines in the study of communication. On the theoretical and epistemological level, I have first tried to establish the difficulty of accounting for ideology within a pragmatic perspective and then reviewed some of the major criticisms CDA faced in the past, and which have still, to my knowledge, not been settled. I concluded that the multidisciplinary heterogeneity of the research programme faces serious challenges; what is more a perhaps unnecessary terminological complexity turns out to be somewhat confusing, since we are left with an unclear idea of what the object of study of CDA is (i.e. discourse and its many acceptations). I have concluded my review of CDA with a few epistemological observations linked to the scientific status of the research programme.

I will therefore not pursue my analysis of manipulative communication within the framework of CDA. However, as mentioned earlier (see §II.1.1.2.3), I do not reject CDA altogether: I still value its socio-political significance and the tremendous effort it puts in practical analysis of text and talk. Even if I do not endorse its theoretical underpinnings and believe that heavy epistemological work is required to make the practice scientifically consistent, its collection of naturally-occurring data is still of great heuristic value.

II.1.2. ARGUMENTATION THEORY

Among contemporary scientific traditions which can to a certain extent be taken to fall within discourse analysis, as their object of study is the argumentative articulation between verbal premises and conclusions, and which have addressed more or less directly – or which can tentatively be taken to be equipped to deal with – issues of manipulative and deceptive communication we find the study of argumentation. In the following section I give an overview of the developments of the discipline from its origin in Greek Antiquity to the latest contemporary mainstream theories, in order to assess the input the present thesis can gain from the study of argumentation.

II.1.2.1. Why argumentation studies are relevant to an account of manipulative and uncooperative communication

Broadly speaking, the study of argumentation is concerned with viable criteria of justification. The overarching goal of a theory of argumentation is thus to provide some kind of procedure allowing oneself to judge whether the conclusion of a combination of propositions (premises) into an argument is warranted or not. In other words, the prime concern of argumentation theory is argument validity. Yet, as a way of overcoming this quite 'atomic' formal level of inquiry – i.e., in order to go beyond a perspective which limits its scope onto formal relationships between propositions, which are usually envisaged in formal logic approaches in the absence of any external constraints –, the study of argumentation incorporated concerns which are emblematic of a construal of argumentation as a social activity, mostly from the 1950s on,[108] so much so that nowadays dialectics has become the dominant paradigms in argumentation studies.

The first aspect of argumentation studies of interest to research on manipulation, and which is the object of the earliest developments of the discipline – i.e. logic –, is the idea that we can isolate formal criteria allowing us to decide whether an argument is valid or not, and, by an extension that most people do not question, it usually follows that this should also tell us whether we should believe the conclusion yielded by the premises or not, whereby if the argument is valid, or sound, it should *ipso facto* lead its audience to believe in the truth of the conclusion (that is, to be convinced). Manipulation is about getting people to accept propositions whose acceptability is problematic by making sure that this is kept concealed from the audience.

[108] The 1950's is the decade which in which the ground-breaking works by Perelman & Olbrechts-Tyteca (1958) and Toulmin (1958) were published. These laid the grounds for an informal shift in argumentation studies.

Logic could thus turn out to provide valuable insights in the analysis of why certain conclusions should not be derived from specific combinations of premises.

In particular, what is remarkable about formal logic is that by providing criteria of sound argumentation it supplies at the same time criteria of fallaciousness. Fallacies may be used deliberately in manipulative endeavours, by endowing an argument with an appearance of validity in order to make an error in reasoning less detectable. With this in mind, any account telling us what a fallacy amounts to is in principle of great value for an analysis of manipulative scenarios in which fallacious reasoning is at work. Of course, many nuances must be adduced within this perspective (see §III.5.1. on fallacies), but it is important first to establish that formal logic, as it enables the analyst to discriminate sound from unsound arguments, provides solid grounds for a formal analysis of manipulation (notwithstanding its limits, which I will evoke further along).

Contemporary developments of argumentation theory, notably those who explore a dialectical direction, may also provide valuable insights for the analysis of manipulation. The dominant conception of argumentation within those traditions stipulates that arguing is foremost a social activity in which the proponent of a standpoint aims at convincing a disagreeing audience (the opponent, or antagonist) of the validity of said standpoint. As a consequence, there is no arguing without prior disagreement, and the need for providing reasons motivating the acceptability of a standpoint is taken to constitute the starting point of any argumentative exchange. Within these approaches, the act of arguing is analysed as a matter of conducting oneself in a way which should lead to the resolution of the difference of opinion, in addition to being a matter of reasoning properly (cf. van Eemeren & Grootendorst 2004). Accordingly, the norms of argument soundness are not solely dependent on formal logic factors; the account also incorporates the dimension of problem-solving ability which should characterise any attempt to convince. Hence, in addition to internal norms of conceptual reasoning such as those we find in formal logic, these accounts develop external norms stipulating what kind of argumentative moves are allowed in an argumentative procedure.[109]

In theories of argumentation of dialectical obedience, fallacies are thus analysed not only in terms of logical (in)consistency, but also according to the extent to which they depart from praxeological standards of argumentation which are usually defined in terms of critical reasonableness (cf. van Eemeren & Grootendorst 2004: 131-134). In principle, adding this type of component to a theory of argumentation makes it possible to deal with naturally occurring

[109] See §II.1.2.3.3.2. for an illustration of this kind of model (cf. Pragma-dialectics, van Eemeren & Grootendorst 2004).

arguments but also with occurrences which display logical flaws while still fulfilling argumentative purposes in an acceptable way. As such, a dialectic model of argumentation may cover more ground than a logical account; as far as manipulation is concerned, it could turn out to be a promising asset, as many cases of deceptive communication are sound from a formal logic point of view.

Pragma-dialectics in particular is interesting as its most recent version also incorporates a rhetorical dimension (cf. van Eemeren 2009), allowing the theory to deal with argumentative occurrences in which the goal of convincing the audience – or to 'win' the argument – competes with the goal of sticking to an ideal of critical reasonableness. It is typically in these types of argumentative exchanges that "strategic manoeuvring" (i.e., the latitude a speaker has at her disposal to enforce her rhetorical aims without compromising the dialectical ideal towards which sound argumentation must tend) may derail and result in the production of fallacious argumentation, which in turn might illegitimately mislead the audience. Few models of argumentation combine a rhetorical component with a dialectical one: in this respect, Pragma-dialectics will receive special attention in what follows, as I will try to assess how such an input can benefit a pragmatic account of manipulation; interestingly, Pragma-dialectics offers different types of norms of argument acceptability we can choose from (from rhetorical to dialectical norms and even logical norms), and for this reason it will be useful to expose them further along in order to assess which of these might be incorporated in the type of account of manipulative communication I am after.

Contemporary approaches to argumentation supply models of argumentative strategies. As such, they have been particularly useful to discourse analysis, a discipline concerned with the structure and functions of discourses, usually within political or social configurations. Argumentation is hence typically one of the facets of discourse that scholars will want to analyse so as to gain access to the effect that it may have on people in terms of conviction or persuasion.[110] Moreover, the use of argumentative norms in principle enables us to identify fallacies; their presence might in turn be questioned from the point of view of their intentionality and constitute a first clue in forming a judgement about a text's manipulative potential. It is also the goal of this section on argumentation theory to assess to what extent available models of argumentation may contribute to a better understanding of phenomena of deception.

[110] See e.g. Ietcu-Fairclough (2008), Reisigl & Wodak (2003) for representative exploitations of argumentation theory within CDA and Amossy (2008) for an illustrative overview of the usefulness of argumentation from the perspective of the francophone tradition of discourse analysis.

II.1.2.2. Formal logic and argumentation

II.1.2.2.1. Syllogisms and formal fallacies

Historically, the origins of the study of argumentation correlate to the emergence of philosophy in Ancient Greece. Van Eemeren *et* al. (1996) note that before that, metaphysical and philosophical questions were traditionally addressed following mythological considerations. Yet, it is usually considered that the quest for alternative (rational) explanations of the world, precisely initiated by Greek philosophers, gave at the same time birth to a novel and authentic concern for justification. Indeed, if a new explanation was to be advanced, then the criteria according to which this explanation could be deemed valid had also to be formulated and evaluated. Thus, out of the need for motivated, coherent and rational explanations of the world emerged the need of establishing stable criteria to evaluate whether a given justification was acceptable or not.

Aristotle (384 BC – 322 BC) was the first to lay the grounds of a formal system meant to guarantee the validity of reasoning assessed in the form of an articulation of propositions; his system of logic, which is still held as a reference frame in contemporary approaches to formal logic, describes the nature and structure of sound argumentation and thus develops a method to evaluate whether a combination of propositions into premises guarantees the truth of a third proposition (the conclusion) derived from the first two. The object he set out to account for was demonstration, that is, a failsafe type of reasoning stipulating that granted the truth of premises, the truth of the conclusion necessarily follows.

We owe him the term *syllogism*, which he defined as a "discourse in which, certain things being stated, something other than what is stated follows of necessity from their being so" (Aristotle, *Prior analytics*, book 1, part 1). A syllogism, of which the deductive demonstration is the canonical instance, stipulates that if the premises are true, then the conclusion necessarily follows, meaning that i) the conclusion cannot be false if the premises are true, and ii) a true conclusion cannot follow from false premises. Overall, the syllogism asserts and ensures the transferability of truth from the premises to the conclusion.

The components of an argument, i.e. the propositions (or statements), are described as consisting of a predicate and a subject, the latter being what is talked about, the former being what is said about the latter. These propositions usually denote relationships of inclusion/exclusion between predicates and subject, which are conceived as sets of objects. Propositions can thus be universal (in case it predicates something of all members of the denoted set) or particular (in case it predicates something about only a limited amount of members of the denoted set). So if one of the premises of the argument is "All men are mortal", I am establishing in that premise that within the set of the objects denoted by the predicate 'being

mortal', we find all (in this sense, 'universally' all) members of the set of objects denoted by the subject 'men'. The conclusion of a syllogism, hence, does no more than a synthesis of the two inclusion/exclusion relationships denoted in the two preceding premises. That is the reason why Aristotle defines a syllogism as a somewhat autonomous piece of reasoning: once the premises have been laid, nothing more than what is contained in those premises is necessary for the conclusion to be drawn.[111]

What formal logic advocates is that provided the premises are true, but also under the condition that the conclusion synthesises the relationships of inclusion and exclusion asserted in the premises, the conclusion cannot be false. Invalid arguments, thus, are those arguments in which the truth of the conclusion does not follow from the truth of the premises. A syllogism does not predicate the absolute truth of the conclusion, but subordinates the necessity of the conclusion's truth to the truth of the premises. As a consequence, what matters in a syllogism is not the notion of truth itself (as can be assessed by checking with actual states of affairs), but a relationship holding between the truth of premises and that of a conclusion which are combined in the argumentative chain.

One of the interesting points about formal logic is that it provides clear means of spotting invalid arguments, and this is especially helpful when the appearance of the argument resembles that of a syllogism. Take for instance an argument such as (53):

(53) a. Some immigrants are inhabitants of the Bronx

b. Some inhabitants of the Bronx are delinquents

c. Therefore some immigrants are delinquents

While it could be argued that the conclusion (53c) may be true (as it may correspond to an actual state of affairs), it cannot be claimed that (53c) is true on the grounds of the combination of (53a) and (53b). This is perhaps not clearly visible, so in order to show how this argument is fallacious, let us replace the variables by more familiar ones:

(54) a. Some birds are inhabitants of the forest

[111] In this respect, some have been tempted to draw a correspondence between syllogisms and the fallacy of *begging the question*. This fallacy describes arguments in which an occurrence of the conclusion is contained within the premises (thus proving circular). To the extent that the conclusion of a syllogism does not adduce new information, some are tempted to contend that syllogisms are instances of the fallacy of begging the question, or at best trivial arguments. This is roughly what John Stuart Mill contended in order to "destroy deduction as a source of knowledge" (Hamblin 1970: 246). See Hamblin (1970: 246-247) for a refutation of Mill's stance, and Iacona & Marconi (2005) for an informed account of the fallacy of begging the question.

b. Some inhabitants of the forest are rodents

c. Therefore some birds are rodents

Here the falsity of the conclusion (54c) is self-evident, because the two classes denoted are mutually exclusive. But what is more interesting on a formal level is that the two relationships of inclusion denoted in the premises cannot constitute conclusive evidence to derive (54c) on the basis of (54a) and (54b). In (54a), a portion of the set of birds is said to intersect with a portion of the set of forest inhabitants. In (54b), a portion of the set of forest inhabitants is said to intersect with a portion of the set of rodents. But nowhere is it said that the two portions of the set of forest inhabitants denoted in (54a) and (54b) coincide, and that is precisely why (54c) cannot be asserted on the sole grounds of (54a) and (54b). The same applies for the argument proposed in (53); it is fallacious because it assumes that the portions of inhabitants of the Bronx denoted in (53a) and (53b) are the same, while there are no formal grounds to assert that, should we have no other information than the one contained in these two premises.

The reason I proposed this example is to show that basic formal logic considerations such as syllogistic relationships of inclusion between sets of objects of the world can prove extremely useful when we are confronted to arguments such as (53), which look very much like plain deductions. In particular, I am thinking that if found in sensitive contexts, such as political ones, such arguments may be misleading to the extent that an addressee might completely bypass the logical flaw by failing to notice that the portions of denoted sets do not coincide. This logical error may have dramatic consequences on the social level: when such a fallacy is used to promote xenophobic ideologies (or any type of discriminative ideology for that matter), granted that its audience fails to notice the error, chances are that the ideologically loaded content will go through without being questioned, thereby leading to potentially harmful behaviour. Formal logic provides argumentative templates, in the form of structures of premises and conclusions; in order to evaluate whether an naturally occurring argument is valid or not, all we have to do is thus to remove its content by replacing it with variables, and compare it with formally valid inference schemes. If the comparison delivers a match, the argument will be deemed valid.

Formal logic would thus constitute a powerful filter to critically test whatever discourse one is processing; not powerful in the sense that it would be a 100% failsafe filter,[112] but powerful in the sense that if applicable to a specific argumentative chain (such as (53)), it has the power to establish its fallaciousness, which is in turn the first step in questioning the speaker's motivation and ultimately gathering evidence to spot a deceptive intent (should there be one).

[112] We will see shortly (see §II.1.2.2.2.) some limitations of logic.

II.1.2.2.2. Limitations of formal logic for an account of argumentation and manipulation: Issues of meaning

If logic may supply useful tools to analyse potential formal flaws in arguments by systematically prompting us to look for logical contradictions and categorical mistakes, the import a theory of manipulative and uncooperative communication can gain from logic is limited for a number of reasons.

First of all, an obvious remark is that arguments in ordinary conversation seldom come in the form of canonical syllogisms, and that assessing 'real life' argumentative exchanges almost always requires *a posteriori* reconstructions, even more so when the argument under examination is enthymematic – i.e., in Quintilian's sense, arguments where either a premise or the conclusion are left implicit. But even if in principle one may reconstruct arguments in a syllogistic form, more often than not this task turns out to be very difficult and the complexity of naturally occurring arguments may resist a formal translation. In the case of enthymemes, for instance, we may be hard-pressed to establish with full certainty that the missing premise we identified is indeed the one the original speaker was committed to as she was uttering her argument, even if our interpretative intuitions are usually quite good. Nevertheless, this is potentially of little importance for formal logic, because, just like formal semantics, it works with abstract and decontextualised representations, at least in what regards propositional and predicate calculus. As Cann puts it, "semantics is the study of meaning abstracted away from those aspects that are derived from the intentions of speakers, their psychological states and the socio-cultural aspects of the context in which their utterances are made" (Cann 1993: 1), and in this sense takes linguistic exchanges and their meaning as objective facts. However, as noted by MacPherson (1999), intensional logic does take into account intentions, meaning and reference (just as dynamic models in semantics like Kamp's and Asher's, which deal with context)[113]; therefore, the critiques dealt with here will concern mostly predicate and propositional calculus, which share with formal semantics the properties exposed in Cann's quote.

Axiomatically, formal logic will not delve into mental states or other external parameters such as socio-cultural settings, but will privilege formal issues of argumentative structure; nonetheless, socio-cultural aspects of communication as well as pragmatic ones do matter and are of great importance when comes the need to assess not only argumentative sequences but also their effect on particular audiences (see Perelman & Olbrechts-Tyteca (1958), van Eemeren & Grootendorst (2004)). This is a first indication that in order to account for argumentation as it occurs in spontaneous and natural conversations (beyond issues of form and thus including the

[113] Saussure, personal communication.

meaning dimension of arguments), one might need to go beyond matters of form and question how the properties of the form are exploited by an individual engaged in a communicative exchange. In principle, the type of formal logic that deals with norms of inference validity is only concerned with structure and relationships (inclusion, exclusion, implication, transferability of truth) between propositional material, independently of their content and of their (broad) context of occurrence. That is precisely why propositional and predicate calculus isolated a limited set of operators (such as conjunction, disjunction, negation, implication – and quantifiers in what regards predicate calculus), but use variables to deal with contents. No matter how you decide to saturate the variables of a logical formula (i.e., whatever referents you opt for), as long as its syntax conforms to the syntactical rules of the system you work within, the formulas you get will be well-formed, and as long as their combination observes the logical rules of the system, they will be valid. Just like mathematical algebra, formal logic aims at assessing whether a combination of variables according to specific operations can be deemed valid, regardless of the actual content one may want to put into the variables and regardless of their context of occurrence.

Because it leaves aside propositional contents by replacing them with variables, predicate and propositional logic are equipped to handle some surprising things, were we to consider them outside a logical framework; for instance, under these requirements, example like (55) or (56) will be deemed logically valid:

(55) a. All confolters are malandreams

b. The ropartling is a confolter

c. Therefore the ropartling is a malandream

(56) a. All ten-legged creatures have wings

b. All spiders have ten legs

c. Therefore all spiders have wings (Copi & Cohen 1994: 62)

Despite its meaninglessness (given that its subjects and predicates are invented words which do not mean anything in English), (55) is a typical *modus ponens* and there is consequently nothing wrong with it from the point of view of its structure, that is, from a purely formal perspective. Should the entities referred to exist in the real world, the argument would be valid and furthermore meaningful, but the question of its meaningfulness does not arise since those entities do not exist in our world. The aspect of formal logic accounts of argumentation I want to stress here is that their concern about form may fall short of providing some essential tools for the analysis of so-called 'real' argumentation, due to its neglecting contents and contexts of propositions. Yet, it must be said that this is not a problem of formal logic (i.e. predicate and

propositional calculus) itself: formal logic was never designed to account for such things, and accusing it of such a limitation is pretty much a straw man fallacy, as the accusation would only hold if their avowed goal was to provide a comprehensive analysis of argumentation which would take into account all of its facets, be they formal or otherwise. But such is not the goal of formal logic, whose business is restricted to the determination of what is a valid inference, so the reservation we could formulate here is not about the (presumed) intrinsic limits of logic, but about the limits of logic *if we want it to fulfil the goal of assessing phenomena of conviction, persuasion and deception*. It is only if we have in mind the goal of understanding how argumentation works globally, from the formal features of the piece of reasoning involved to the arguer's and the audience's behaviour in the argumentative exchange, that formal logic will appear to be unsuited. But this is no argument to dismiss it altogether; it is rather an indication that it should be combined with other types of accounts to deal with the bigger picture.[114]

(56) is also interesting because it is logically valid yet unacceptable should we go by semantic or pragmatic standards. (56a) is a false premise, as it is not true in our world that all ten-legged creatures have wings. From this follows that (56c) will also be false. What is interesting is that the form of the argument is valid: if we replace the valuations by variables, we get the exact same form as the *modus ponens*, i.e. a deductive inference ('All A's are B's; All C's are A's; Therefore, all C's are B's'). Again, formal logic does not care about the actual truth of premises and conclusions, but will only be concerned with the fact that *if* the premises are true, then the conclusion will follow. Example (56) shows precisely that arguments with false premises and a false conclusion may still be valid.

This last point about the importance of meaning in argumentation can also be illustrated in the following way; let us take the example of a tautological argument (57a), for which logic may have little to say while other disciplines may provide insightful comments:

(57) a. I don't support Bush, because Bush is Bush

b. Bush has done some terrible things in the past

c. Like his father George W. Bush has done some terrible things

The argument introduced in the second clause to back up the first clause is tautological since it is of the form 'A is A'. From a logical point of view, there is nothing wrong with this formula – even if it is trivial – because it is always true, regardless of its possible interpretations, that is,

[114] And in fact, no approach to argumentation – and it is also plausible to consider that no scientific approach – can do without some founding principles of logic, such as the principle of non-contradiction and deductive laws of implication.

regardless of the valuation one decides to replace A with. When the tautology used in (57a) functions as a premise to support the conclusion that I do not support Bush, little can be objected from a logical point of view, because the argument turns out to be valid. Now, of course, when processing (57a), we are led to construct a more substantial, or relevant, meaning of the second clause, perhaps something like (57b) or (57c). These two interpretations of the second clause of the argument are much more relevant than what the tautology literally encodes, and they represent two plausible outcomes of the interpretative process hearers would go through in order to pragmatically resolve the tautology. What this example shows is that if we want to address issues of argument processing from the perspective of meaning and interpretation of the actual content of arguments, logic may only have to offer limited insights into these phenomena. I suggest that this is because the issue of (pragmatic) meaning is not the object of formal logic.

If we pursue our inquiry about the cognitive implications of argument processing, a few questions can be raised regarding the plausibility of formal logic as a system meant to map natural norms of rationality. Logic is, according to Lepage, meant to "characterise in a general way the notion of *valid inference*" (Lepage 1991: 1, author's italics, my translation). Yet, nothing in this definition tells us that 'valid inference' should be understood as '*natural* valid inference'. As a matter of convenience, of course, it could be assumed that the valid inference schemes formal logic describes are default and spontaneous schemes of natural reasoning, that is, that these inferences have psychological reality. But this calls two interrogations: first, pairing this with the idea that this is how our reasoning abilities *naturally* work is questionable for a number of reasons, mainly stemming from literature on reasoning errors and cognitive illusions.[115] Second, this brings up the idea that perhaps conviction, as the ultimate goal of argumentation, is not solely dependent on formal validity. I will address this second issue in the next section through an exposition of the main tenets of informal logic and their arguments directed against the suitability of formal logic principles as argumentative norms before I turn to dealing with the first issue towards the conclusion of this sub-chapter on argumentation theory.

[115] See Pohl (2004) and *infra* (§II.3.1.).

II.1.2.3. Informal logic

II.1.2.3.1. Precursors: Toulmin and Perelman & Olbrechts-Tyteca[116]

Informal logic was born partly in reaction to the limits formal logic encountered when attempting to deal with the evaluation of the validity of naturally occurring arguments which did not seem to conform to logical standards. As a precursor of the informal turn in argumentation studies, Toulmin's *Uses of argument* (1958) contributed to laying the grounds for the development of informal logic. Instead of limiting his account to formal norms of inference validity, he tried to broaden the scope of argumentation theory through a model designed to deal with justification more broadly conceived. This way, any piece of reasoning could in principle be dealt with within the system, and in particular those arguments which do not comply with formal standards of validity but that we do nonetheless find in everyday argumentative conversation.[117] Arguing against an exclusive and universalistic norm of formal validity, Toulmin contended that field-dependent (and thus relative) norms are also relevant to the evaluation of arguments. In particular, he held that a formalisation of *practical reasoning* (as opposed to formal reasoning) was possible through a stable abstract argumentative pattern, which he formalised as follows: a claim (C) is supported by data (D), whose relevance to the point is guaranteed by a warrant (W) that implicitly or explicitly constitutes the justification of why the data can be used to back up the claim. If the warrant is called into question, then a backing (B) may be supplied for further justificatory reasons. In case a rebuttal (R) emerges, we can also specify or weaken the claim with a qualifier (Q). It is noteworthy that C, W and B deal with the content of the argument, not with its form, i.e. they are steps whose acceptability rests on that of the denoted objects (and hence the field-dependency of the model). At the same time, however, this procedure is also characterised by Toulmin as a *field-invariant* procedure, since he assumes that all arguments follow it. As such, it is in principle also compatible with different types of reasoning. In (64) I reproduce Toulmin's classical example for illustrative purposes:

(58) Harry was born in Bermuda (D) since a man born in Bermuda will generally be a British subject (W) on account of the following statuses and other legal provisions (B); so, presumably (Q), unless both his parents were aliens/he has become a naturalised American (R), Harry is a British subject (C). (adapted from Toulmin 1958: 97)

---

[116] Parts of this section are based on van Eemeren *et al.*'s (1996) thorough exposition of the main approaches in the history of argumentation.

[117] This is crucial in identifying the conditions under which sometimes we end up convinced for everything but formal reasons.

Toulmin's approach is normative in a procedural sense, not in a qualitative or conceptual sense. He postulates, following his observations on legal jurisprudence, that all arguments – analytic (logical entailments) or substantial arguments (non-logical entailments which are nevertheless validated) alike – share a similar general pattern of justification.

Now, the above procedure does not make argument soundness depend on whether they follow this procedural, structural norm. Rather, acceptability will rely on a criterion of *field-dependency* based on the content of the argument and its adequacy to the field it belongs to. Toulmin's primary contribution is that of addressing argumentation theory from an epistemological stance: he is interested in the way people come to know new information. In particular, he contends that the fact that the same argument can be convincing to certain audiences but not to others disqualifies a formal account. Argument soundness must therefore be evaluated against its adequacy to the audience and to the issue at stake, which lies in the degree to which the warrant is made acceptable by the backing. In the end, this conception turns out to be relativistic, since it boils down to a judge to evaluate whether the warrant is indeed supported by the backing. As a consequence, argument validity rests on the evaluation of a judge which belongs to the field related to the issue at stake (for instance, an acceptable judge for an argument advanced to choose one particular computer over another would be a computer technician). This entails that there can be as many sound arguments as there are acceptable judges. Validity is thus field-dependent-bound.

The model raises several issues, notably terminological ones, as noted by van Eemeren *et al* (1996: 155). The distinction between warrant and backing remains problematic as well, because the backing could be construed as a sub-argumentation relating to the warrant, opening the path to complex argumentations, which Toulmin does not address directly. But above all, the oscillation between normative and non-normative considerations may turn out to be a bigger issue. The procedure arguments follow is deemed to be universal, and as such it is normative. However, such normativity does not apply to argument soundness, because arguments are field-dependent. From a cognitive point of view, this might also constitute a serious shortcoming. On the one hand, the model lacks a clear justification of why this procedure should correspond to what the human mind actually does when processing argumentation, and on the other hand, the fact that the model turns out to be relativistic is incompatible with a cognitive naturalistic model of argumentation, if only because no stable and clear-cut relationships of cause and effect (with respect to the conditions under which an argument can be said to be valid) can be identified. And yet, we feel more can be said about the way we process information – in particular arguments – from a cognitive perspective; moreover, if, as I think is the case, we are able to document recurring if not systematic biases in information-processing that correspond to

documented fallacious argumentative patterns such as the *ad* arguments, it would support the idea that mechanisms of information-processing can contribute to a better understanding of argumentatively flawed communication, and possibly manipulation. For these reasons and considering the limitations of this theory, which cannot straightforwardly account for the problems I have in mind, I will not make use of Toulmin's model of argumentation in the rest of this thesis; however, his contribution was important mentioning, to the extent that it historically contributed to the establishment of informal logic as a mainstream theory of argumentation.

We owe the second cornerstone of the informal turn in argumentation studies to Chaim Perelman and Lucie Olbrechts-Tyteca's *Nouvelle rhétorique*, also published in 1958. Its authors made two simple observations about legal argumentation, as witnessed in law courts, prior to developing an alternative model: i) certain statements escape empirical confirmation or formal proofs (for instance moral or ethical judgements), and ii) one does not need to argue in favour of a statement everyone accepts as evident (an evident truth needs not be debated), which calls into question the notions of evidence and truth so central to formal logic. It emerged from these observations that the very nature of argumentation is highly conflictual, in that the point of arguing is not to argue about that which is true, but to argue about that which is plausible (this requiring that at least one of the arguers does not think that what is being discussed is the case). The real goal of argumentation is consequently to gain an individual's adherence to a thesis, not to establish, in absolute terms, whether a thesis is true or false. This in turn supposes that argumentation involves at least two participants. According to *La nouvelle rhétorique*, any perspective on argumentation (be it logical or rhetorical) should be augmented with a strong dialectical contribution. On a philosophical level, this conception calls into question the type of rationality against which argumentation was traditionally assessed ('hard' individual rationality).

In *La nouvelle rhétorique*, standards of classic formal logic are disregarded as they are deemed unsuited to answer the novel questions argumentation theory needs to address, and focus is set on the notion of audience, since it constitutes a key parameter in the assessment of sound argumentation. Two observations lead to this idea: i) a single argumentation can be effective for some audience, while not for another (or the same audience at another time), and ii) an individual may very well be convinced *to a certain extent*, which is problematic for an absolutist and universalistic acceptation of the notion of conviction. Therefore, Perelman and Olbrechts-Tyteca posit that knowing the audience is crucial to the success of the argument. The rigidity of formal logic should be sensibly softened in order to account for these observations, and one of the ways of achieving this is by adopting a more relativistic stance consisting of the idea that argumentation has to be accommodated depending on the audience it is directed to. Perelman

and Olbrechts-Tyteca thus suggested that in order for an argument to be effective, it must take into account (and comply with) the values, desires, beliefs (i.e. the mental states) of the audience it is intended to convince. Two types of audience are distinguished: i) the universal audience (an ideal complying with logical argumentation), and ii) a particular audience constituted by the set of individuals the speaker wishes to convince (towards the mental states of which the arguments should be tailored). In this respect, an audience is envisaged as a speaker's sociological and psychological construction.

La nouvelle rhétorique elaborates a procedure involving discursive means geared to modify the epistemic dispositions of an addressee. This argumentative procedure is twofold: first the speaker forms his representation of the audience and shapes his discourse (by carefully selecting the premises he deems acceptable to the audience),[118] and then goes public according to these representations. Successful argumentation, under this perspective, always supposes that the speaker needs specific knowledge about his audience, and, ideally, involves some kind of tentative agreement between the discourse and the people to whom it is directed.

On a more technical note, Perelman and Olbrechts-Tyteca distinguish several types of argument schemes (*loci*) on the basis of a criterion of association/dissociation (i.e. whether the argument brings information together that was previously separate, or, on the contrary, disintegrates units of information in smaller, separate, parts). However, this criterion turns out to be strictly typological, since in the end what will determine the argument's soundness is its acceptance by a given audience; it boils down to establishing that certain argumentation schemes will be instrumental to convince certain audiences who are receptive to such schemes, while other audiences will remain unconvinced by them.

Several problems emerge from this conception of argumentation.[119] First of all, similarly to Toulmin's model, the relativistic aspect of the model is problematic with respect to the notion of rationality: if the researcher does not rely on some kind of stable set of standards, but rather postulates a notion of argument validity particular to certain audiences, then the outputs yielded by the model are infinite in theory (or at least as numerous as the number of possible audiences), which constitutes an explanatory shortcoming in the theory. Second – and consequently –, the practical application of the model is problematic in that it does not allow analysts to reach a unified result, as these will be as varied as the audiences potentially involved

[118] Premises can range from facts to presumptions, values, hierarchical schemes or *loci*. What is more, we find in this conception the notion of selection of premises, which will be something crucial in the account of argument processing I will elaborate in chapter III.

[119] See van Eemeren *et al* (1996: 119-128) for a thorough discussion.

in a given argumentation. This makes the theory subjective, a mere consequence of the lack of normative groundings in the model; it is actually one of the major critiques van Eemeren *et al*. (1996) address to Perelman and Olbrechts-Tyteca. Even if the notion of universal audience could be taken as one norm against which argumentation could be evaluated, its formulation is too vague and we cannot expect it to provide solid scientific grounding as a norm. *La nouvelle rhétorique* addresses *effective* argumentation, and therefore can be taken to focus on persuasion, but it cannot supply testable or verifiable predictions; no conditions of success are proposed otherwise than through the reliance on the relative notion of audience, which, as has been pointed out by numerous scholars, is perhaps the major downside of this approach.

The precursors of informal logic have inaugurated another way of approaching the question of inference validity, by refusing to reduce it to the notion of demonstration and by taking into account 'real-life' argumentation. In doing so, they acknowledge that logic is limited to the explanation of demonstration (usually by considering deduction and formal proofs in which conclusions are drawn by necessity) and argue in favour of a softer conception of argumentation that goes beyond the formal paradigm. As such, their enterprise constitutes a step forward in accounting for the reasons people end up convinced of certain propositions, as the latter cannot be condensed into the sole deductive system. Perelman & Olbrechts-Tyteca define the study of argumentation as "the study of the discursive techniques allowing [the arguer] to provoke or increase the minds' adhesion to theses that are presented before their assent" (Perelman & Olbrechts-Tyteca 2008 [1958]: 5, my translation). From this definition, it is clear that the object of a theory of argumentation is not the same as the object of a theory of formal logic. Where the latter aims for necessity, absolute validity, from which you are entitled to derive truth granted the truth of the premises, the former makes quite a lot of room for plausibility and issues of consent (or "adhesion"). This is why informal logic opens up in principle a promising direction of research. Yet, we have seen that the nature of these models leaves little room for cognitive considerations; accordingly, they are under-equipped as it were to fully address the issue of adhesion and consent, which seems to require the import of psychological and cognitive insights in order to gain explanatory plausibility.

II.1.2.3.2. Fallacy theory

The study of argumentation within informal logic can take different orientations, according to the reasoning patterns they opt for; while some argue that all argumentation may be reduced to a deductive model (the argument being easily reconstructed by adding missing premises or conclusions), others (e.g. Walton 1987, 2006, 2008) claim that the tension between deductive, inductive, abductive or conductive reasoning cannot be resolved by privileging one pattern over

another. One of the ways of overcoming the difficulty has been to tackle argumentation indirectly, through the study of argumentative errors, or fallacies. Taking as a starting point Aristotle's *Sophistical Refutations*, these approaches elaborated over time what became known as the "standard treatment of fallacies". Hamblin, who was one of the first to provide a state of the art of this standard treatment, recalls that a fallacious argument is standardly conceived as "one that *seems to be valid* but *is not so*" (Hamblin 1970: 12). The definition leaves room for two options in the study of fallacies: either we focus on that which makes them invalid or on that which makes them seem valid. However, Hamblin points to the failure of fallacy theory, in that "we have no theory of fallacy at all, in the sense in which we have theories of correct reasoning or inference" (Hamblin 1970: 11). This is perhaps because fallacies have traditionally been construed as a tool to address informal reasoning, and consequently studied descriptively and typologically. In fact, most accounts, from Locke onwards, try to identify different types of fallacies (for instance as 'ad- arguments') and to describe their conceptual structure and the way they relate to the issue at stake.

Hamblin set out to provide a model which would account for fallacies in argumentative dialogues, and proved extremely influential for argumentation scholars in the end of the 20th century. Notably, he was the first to speak of *formal dialectics* as "the setting up of simple systems of precise but not necessarily realistic rules, and the plotting of the properties of dialogues that might be played out in accordance with them" (Hamblin 1970: 256). Similarly to Lorenzen's (1969, 1987), Hamblin's project was to envisage argumentation as a practice, away from the abstract conception formal logic had defended for centuries.

Among the modern practitioners of fallacy theory, Woods and Walton[120] published a collection of work on fallacies throughout the last two decades, either individually or in collaboration. Their prime concern is to assess the functioning of fallacies in actual conversation. This resulted in a comprehensive descriptive and typological endeavour, which has dealt with more than twenty instances of fallacies and more than sixty-five argument schemes. One of the peculiarities of their work is the fact that there is no clear-cut position as to the nature (as exclusively formal or dialectical) of fallacies. This probably stems from the fact that their assessment relies sometimes on plain logical factors, sometimes on dialectical grounds; in other words, they provide *ad hoc* accounts. Underlying the analysis is the assumption that fallacies are addressable according to different logical systems (or reasoning patterns), and therefore that no unified framework is conceivable in terms of the classical distinction between formal logic, dialectic and

[120] See for instance Woods & Walton (1989), Woods (1995), Walton (1987, 1989, 2006).

rhetoric. Walton's is therefore a pluralistic account, where fallacies are ascribed different degrees of formality, which turns out to be compatible with the informal project.

Walton's model has gradually evolved into a more dialectical approach, with its reliance on the notion of dialogue types. Argumentation is conceived as a "social and verbal means of trying to resolve, or at least contend with, a conflict or difference that has arisen between two parties engaged in a dialogue" (Walton 1990: 411). Fallacies, as a consequence, will be evaluated in terms of their (in)appropriateness to the dialogue in which they occur, and not in terms of their form. One single argumentative move, for instance a threat, can therefore be admitted in an eristic dialogue (i.e. interpersonal conflict), but not in a persuasive dialogue, since it is expected and allowed (by people's knowledge of dialogue types) in the former, while it is not in the latter. This builds an intimate link between the notion of fallacy and that of legitimacy. Fallacies are illegitimate moves, in turn giving evidence of Walton's normative approach to argumentation.

On a more technical note, Walton objects to the 'belief-desire-intention' model (cf. Bratman 1989) on the grounds that beliefs, desires and intentions are private, and not accessible in their natural form. He claims that as he himself is not even aware of all his beliefs and intentions, their study is *ipso facto* compromised, somehow subscribing to the classical behaviourist position which holds that the mind is a black box about which little can be said. According to him, any such account will lack plausibility. Instead, he advocates the *commitment model*. Commitments are public, because of their explicit nature. In principle, one can always assess whether a speaker has gone on record with respect to a certain statement. This is crucial in argumentation, especially if conceived as a means of resolving a difference of opinion; if we are able to ascribe public endorsements of statements, then we will tentatively be able to work out a solution in a way that leaves little room for speculation (since we build on data which has been publicly made available).

One of the main critiques that could be addressed to Walton's approach concerns its problematic normative aspect. If we define a fallacy as an illegitimate move, this has to be done with respect to a certain norm. It remains however unclear where this norm lies, in particular whether it is a cognitive, formal or social norm. As his analysis of fallacies oscillates between logical, rhetorical and dialectical frameworks, so do the norms he makes use of.[121] Furthermore, Walton does not justify otherwise than empirically the actual existence of dialogue types, nor does he provide grounds on which such distinctions could be thought of. There seems to be an additional shortcoming with respect to the notion of conviction (or with respect to the reasons why an argument is convincing): it is one thing to describe how an argument works (with reliance on a

[121] See van Eemeren & Houtlosser (2007) for a discussion of Walton's contribution to informal logic.

specific type of reasoning and with respect to a particular dialogue type), but it is quite another to assume that the description actually reflects what happens in people's minds when they process that argument. Once more, just like other informal logic approaches to argumentation I have mentioned, it seems that contemporary fallacy theory faces the objection stipulating that a lack of cognitive insights is detrimental to its account of conviction.

I now turn to exposing a third direction of research, namely dialectics, which is certainly the most influential paradigm in argumentation studies nowadays.

II.1.2.3.3. The dialectical shift

II.1.2.3.3.1. Formal dialectics

Informal logic gave rise to a third strand of research in the study of argumentation, namely a development of the dialectical perspective. The proponents of this theoretical stance conceive of argumentation in teleological terms, as a dialogical exchange between language users meant to resolve a difference of opinion regarding a particular standpoint. It requires systematic submission of statements to doubt, therefore forcing participants to defend them. The argumentative exchange thus consists of a sequence of moves, regulated by implicit norms, in the tradition of discourse analytical approaches building on traditional Searlean and Gricean accounts of human communication. An argument is not reduced to a conceptual articulation to be evaluated according to its acceptability or soundness; it is conceived in praxeological terms as a sequence of actions involving at least two participants. Thus matters of success and efficiency are of prime importance, a position taken as a departure from formal logic considerations, since it is not so much the validity or the plausibility of the argument which matters most, but rather its ability to solve the dispute.

Most of these approaches build at least partly on the theory of *Formal Dialectics*, as exposed by Dutch logicians and philosophers Else Barth and Erik Krabbe, who themselves elaborated their model drawing on Hamblin's but also Lorenzen's theoretical work.[122] The latter, with the Erlangen School, developed a system of dialogue logic in which argument validity, notably postulated as an articulation of propositions around logical connectives, can be established if the participants agree on the same atomic propositions. A logical proof, in these terms, is a dialectical construction which tests standpoints by their confrontation to counter propositions. In this respect, Lorenzen's project is similar to Hamblin's. Barth and Krabbe took this dialectical definition of logical validity and worked it out so as to supply a consistent model of critical dialogues. For them, arguing is a practice between at least two participants (the proponent and

[122] See Barth & Krabbe (1982), Hamblin (1970), and Lorenzen (1969, 1987)

the opponent), destined to settle a dispute concerning a particular claim. The argumentative procedure assigns duties and rights to the participants, thereby involving a normative agenda. The dispute is settled when agreement on a proposition has been reached. It should be noted that in this perspective arguing is conceived as a game in which there is a winner and a loser.

According to Hamblin, when agreement is reached, it is because either the opponent or the proponent has been led to accept his adversary's proposition. This can only be evaluated in light of the commitments participants have publicly taken; if a contradiction between the commitments of one participant is made manifest, he must recognise that his opponent is right. This is how the notion of commitment enters the picture as an empirical method meant to settle the dispute. Barth and Krabbe take the notion one step further and postulate that a commitment is a contract a participant has publicly taken, resulting from a dialogic *exchange*, not from an individual speech act. This, in turns, supports the view that commitments should be mutually recognised by participants in order to be contractual. Barth and Krabbe's model (1982) is extremely constrained, in that there are only three ways of ending a dispute: i) when one of the participants physically leaves, ii) when one of the parties has behaved irrationally (i.e. has infringed the dialectical rules), or iii) when one of the parties can no longer make an admissible move, except by breaking one of its commitments. This view on argumentation shows quite radical and raises issues as to its plausibility (for instance, one would have the intuition that retracting from commitments is still possible in ordinary conversation).

II.1.2.3.3.2. Pragma-Dialectics

II.1.2.3.3.2.1. Theoretical and practical aspects of Pragma-Dialectics

II.1.2.3.3.2.1.1 The standard pragma-dialectical model

Perhaps the most evolved development of formal dialectics is Frans van Eemeren and Rob Grootendorst's Pragma-Dialectical theory of argumentation (henceforth PD).[123] Within PD, argumentation is considered as a "verbal, social, and rational activity aimed at convincing a reasonable critic of the acceptability of a standpoint by putting forward a constellation of propositions justifying or refuting the proposition expressed in the standpoint" (van Eemeren & Grootendorst 2004: 1). The overarching contribution of PD is to consider argumentation as a phenomenon of actual interaction, in addition to a matter of abstract conceptual structures. This benefits the theory by adding contextual data as a parameter of crucial importance. Arguing is therefore an activity undertaken in order to resolve a difference of opinion by advancing

[123] For some of the most important references in Pragma-Dialectics, see van Eemeren & Grootendorst (1984, 1992 and 2004), van Eemeren *et al*. (1996), and van Eemeren (2010) for the most recent version of the Pragma-dialectical theory of argumentation.

propositions sustaining the claim whose acceptability is being questioned. This presupposes first that there are two participants, one of them casting doubt on the acceptability of the other's standpoint.[124] As a result, argumentation is a dialectical process: not only does it involve the participation of two individuals working out the resolution of a dispute, but it also requires systematically submission of statements to doubt, therefore forcing their proponents to defend them.

PD is an ambivalent model: argumentation is conceived of both as a process and as a product: "The term *argumentation* refers at the same time to the process of arguing ('I am about to complete my argumentation') and to its product ('This argumentation is not sound')" (van Eemeren & Grootendorst 2004: 1). As a process, it is conceived as a goal-oriented activity realised by the speaker's performance of speech acts. Being a product, argumentation is the result of this process, i.e., the performance of argumentative speech acts. These considerations pertain to the point of view of argumentative production.

However, PD also envisages the perspective of reception through its analytical application, since it supplies a model of argument evaluation geared to assess the acceptability of arguments:

> "Using the [critical discussion] model as a guide, the reconstruction aims to produce an *analytic overview* of all components of a discourse or text that are pertinent to the resolution of a difference of opinion. Pursuing this aim involves examining exactly which points are at issue, which procedural and material points of departure are chosen, which explicit, implicit, indirect, and unexpressed arguments are advanced, which argument schemes are used in each single argumentation, and how the argumentation that is formed by combining single argumentations is structured." (van Eemeren & Grootendorst 2004: 96)

This way, both production and reception of discourse can be attended to. This is considerably advantageous, as from a linguistic perspective, in principle PD allows addressing issues relative to the speaker's *and* to the hearer's contribution in the argumentative exchange.

PD is based on SAT (following Austin 1962 and Searle 1969) and on Gricean pragmatics, in the sense that the interaction in which argumentation is embedded is taken to observe certain

[124] There need not be two *actual* participants, since one can argue by oneself in order to judge the acceptability of a proposition by casting doubt on it and deciding upon reflection whether it is valid or not. However, casting doubt presupposes both a standpoint and its being called into question, and, by extension, it involves two participants (the one holding the standpoint, and the one calling it into question) – even if one of them remains *virtual*. We find here Perleman & Olbrechts-Tyteca's idea according to which we always argue because the claim under discussion is not self-evident.

standards and conventions: it complies in this respect with an elaborate version of Grice's Cooperation Principle, the "Communication Principle" (See van Eemeren & Grootendorst 2004: 76-77). PD thus regards argumentation as a conventionally regulated social practice. In this respect, it can also claim an affiliation to the field of discourse analysis, where discourses are seen as 'corpus wholes' and studied from the outside, in terms of a compliance of their structure (the dependence relations their constituents have with each other) to conventional standards.

The conceptual core of PD is the ideal model of the *critical discussion*. It is defined as follows:

> "By a *critical* discussion we mean a discussion between a protagonist and an antagonist of a particular standpoint in respect of an expressed opinion, the purpose of the discussion being to establish whether the protagonist's standpoint is defensible against the critical reactions of the antagonist." (van Eemeren & Grootendorst 1984: 17)

The reconstruction of critical discussion produces an *analytic overview* of the discourse, which is a global picture of the argumentation that took place, divided in four stages: *confrontation* (making the standpoint explicit and accepted as a questionable standpoint), *opening* (the participants manifest themselves as parties – protagonist and antagonist – and determine whether there is a common ground to conduct a meaningful exchange), *argumentation* (participants advance arguments in order to overcome doubts regarding the standpoints), and *concluding* stage (the participants establish whether the standpoint has been successfully defended or not). It must be noted that these stages, except the argumentation stage, can remain implicit, but they have to be made explicit afterwards in the reconstruction of the critical discussion when we are to assess the resolution of the difference of opinion (see also van Eeemeren and Grootendorst 2004: 57-62).

Through a critical discussion, which is the ideal format behind any argumentative discussion, participants (protagonist and antagonist) exchange views in order to arrive at a resolution of the dispute by agreeing on the acceptability or unacceptability of the standpoint called into question. This procedure unfolds following – in its standard version (van Eemeren & Grootendorst 2004: 136-157) – a set of fifteen rules "intended to enable language users to conduct themselves as rational discussants". These are also "calculated to prevent anything that might hinder or obstruct the resolution of a dispute" (van Eemeren & Grootendorst 1984: 151). In parallel, there are also ten commandments (van Eemeren & Grootendorst 2004: 190-196) listing the prohibited moves that would be detrimental to the resolution of the dispute and which simplify the fifteen rules by focusing on prohibitions.

From a philosophical perspective, PD considers that successful argumentation requires the observance of standards of critical reasonableness which underlie the rules for a critical

discussion. These standards basically boil down to two characteristics: a sound argumentation should be instrumental to the resolution of the dispute (i.e. it should have problem-solving ability), and it should make use of conventionally accepted moves (either mutually accepted propositions or argumentation patterns).

The analysis of argumentation is carried out by reconstructing the argumentative path undertaken by the participants and by evaluating it with regard to the aforementioned rules and principles. However, this approach to argument validity differs from formal or logical approaches to argumentation, mainly because it is *pragmatic* (where pragmatics denotes, via SAT, a shift from propositional concerns to illocutionary ones, hence proposing a way of dealing with semantic underdeterminacy and integrating contextual considerations) and *dialectic*, since argumentation is thought of as a social activity involving a discussion procedure regulated according to standards of critical reasonableness. As a direct consequence of this type of approach, fallacies are no longer merely conceived as conceptual faults or logical mistakes, but rather as inadequate or forbidden pragmatic moves performed in argumentative discourse by a participant. Classical logic assesses validity relatively to the conceptual structure of argumentation, independently from utterance production, whereas PD does so relatively to the very *performance* of argumentation. In fact, PD extends the conception of fallaciousness to matters of illocutionary inappropriateness, thus avoiding a restriction of its scope to an exclusive assessment of operations concerning propositional contents.

Those speech acts that do not go by the rules of the critical discussion will be discarded as valid arguments insofar as they do not contribute to the resolution of the dispute and will thus be labelled as fallacious. In sum, fallacies are speech acts that violate the rules (i.e., the rules for a critical discussion, though it can be the case that these match speech act felicity conditions).[125] Let us take an example to illustrate this strong claim of PD:

(59) Laszlo's arguments are nonsensical; everybody knows he spent some time in a mental institution some months ago.

(59) can illustrate a violation of Commandment # 6 ("Discussants may not falsely present something as an accepted starting point or falsely deny that something is an accepted starting

[125] For a discussion of the conditions that have to meet for argumentative speech acts to be felicitous, see Chapter 3 of van Eemeren and Grootendorst (1992, especially p.26).

point", cf. van Eemeren & Grootendorst 2004: 193) as follows:[126] if the fact that Laszlo was admitted in a mental institution has not clearly been established beforehand as an accepted premise by the participants, then (59) is fallacious in that it supposes that this has been the case. As we can see, here, the problem does not lie with the content of the utterance nor in the logical articulation between clauses, but within what the speaker *did* by uttering (59).[127]

II.1.2.3.3.2.1.2. Strategic manoeuvring as the incorporation of rhetoric in pragma-dialectics

I have exposed how the dialectical ideal of the critical discussion provides a normative framework for the assessment of argument validity. Yet, its rigidity has been called into question and attacked on the grounds of a simple observation: people in ordinary conversation seldom appear to observe the rules of the critical discussion. For instance, most argumentative exchanges are never settled by an explicit concluding stage, where both participants would agree on a standpoint and acknowledge that the converse position has to be abandoned. It is probably due to such observations that the pragma-dialectical school of argumentation has undertaken a rhetorical shift, which by no means amounts to a dismissal of dialectical ideals altogether.

Rather, in its extended model PD now contends that in addition to sticking to a critical-reasonable posture, arguers try to fulfil rhetorical goals at the same time, and that this can still be accounted for within the pragma-dialectical system. Through the notion of *strategic manoeuvring*, PD nowadays provides a model of how people try to win arguments, without necessarily violating the standards regulating the critical discussion; specifically, and in an effort to conjugate into a same model the different goals arguers may be pursuing as they engage in argumentation, PD considers that "effective persuasion must be disciplined by dialectical rationality" (van Eeremen & Houtlosser 2000: 297). This way, in addition to providing a model of argument validity, PD develops a model of argument effectiveness, whereby issues of

---

[126] It should be also noted that in some contexts, (59) can be purported to carry a fallacy known as *ad hominem attack*. This consists in introducing irrelevant premises about the proponent, in order to make him, and as a consequence his arguments as well, untrustworthy. It usually gets the hearer distracted from checking the validity of the arguments, which is a also fallacious move.

[127] One could object that classical logic already addressed these issues, notably when dealing with fallacies shifting or reversing the burden of proof (that is, making your opponent prove that your own statement is wrong instead of proving yourself that it is right). The contribution of PD, however, is to be sought on the discursive level, to the extent that their study of argumentation broadens its scope to sociological considerations. Argumentation, as a social practice, is analysed as an *event* with social conditions of production, and therefore in relation to its context.

conviction and persuasion become part of the general theory so that the overall account gains plausibility.

Accounting for strategic manoeuvring can easily be done within the dialectical model of the critical discussion, as it is assumed that arguers may manoeuvre strategically in all four of its stages (confrontation, opening, argumentation and conclusion). There are three aspects to strategic manoeuvring: i) the participant who wants to manoeuvre strategically may try to orient the topical potential of the exchange, by for instance getting the conversation to deal only with those particular issues which s/he feels comfortable with, and, more generally, by managing to control the set of contents which will partake in the exchange; ii) strategic manoeuvring may be guided to meet audience demand in all stages, so that every argumentative move, be it formal or semantic, is tailor-made to resonate with the audience's values, preferences, goals, desires and beliefs; iii) presentational devices are the last aspect of strategic manoeuvring: they are effective discursive means of performing moves and in this category we will typically find phenomena linked to the wording of contents, such as stylistic figures which can be used for their effectiveness on audiences.

Strategic manoeuvring can obviously go wrong, or 'derail' according to the official terminology, if the rhetorical manoeuvre supersedes the dialectical effort. If for instance a speaker fails in just trying to suggest the adoption of a particular starting point she feels comfortable with and ends up actually imposing it on her opponent by force (through an argument *ad baculum* for instance), we will say that strategic manoeuvring has derailed, and that a fallacy was committed. This is interesting for the global theory because it indicates that we can also look at fallacies from a rhetorical point of view, rather than limiting ourselves to dialectical considerations. In fact, it seems that whenever a fallacy is committed, this is the case both on the rhetorical and the dialectical level, since "all derailments of strategic manoeuvring are fallacious and all fallacies can be regarded as derailments of strategic manoeuvring" (van Eemeren & Houtlosser 2002: 142).

Now, how do we determine whether strategic manoeuvring has derailed? PD's answer is that the criteria which allow us to classify a particular strategic move as fallacious are contextual, that is, that each activity type in which participants may be engaged has its specific criteria of what we could call 'strategic fallaciousness'. Similarly to Walton's approach to dialogue types, PD advocates that certain moves may be allowed in certain activity types (negotiation, cross-examination, debate, interview, etc.) but forbidden in others. It follows from this that we need to take into account the situated practices in which argumentative discourse unfolds before we can judge the validity or fallaciousness of arguments: "fallacy judgments are in the end always

contextual judgments of specific instances of situated argumentative acting" (van Eemeren & Houtlosser 2002: 142).

The notion of strategic manoeuvring represents PD's way of softening the normative and rigid tone of the standard model, which has been challenged for its idealness, so that the overall theory becomes more flexible by reflecting actual argumentative exchanges. By incorporating a context-dependent component in the analysis, the model strengthens is practical adequacy to what we happen to see in spontaneous and ordinary argumentative exchanges. Strategic manoeuvring is thus to be conceived as a way of relaxing – without abandoning them, of course – the requirements imposed by the criterion of critical reasonableness, as something arguers have at their disposal in order to fulfil rhetorical goals (like winning the argument) without being constrained by the rigidity of a demanding normative system. It should nonetheless be said that nowhere in the pragma-dialectical theory of argumentation is it mentioned that strategic manoeuvring is by definition a hindrance to the pragma-dialectical observance of critical reasonableness: strategic manoeuvring is compatible with the critical discussion and may be perfectly legitimate in conversation. The only situation in which strategic manoeuvring and dialectical requirements deviate into fallaciousness is when strategic manoeuvring derails (i.e., when a given argumentative move either breaks one of the ten commandments or happens to be inappropriate in the activity type in which arguers are engaged).

II.1.2.3.3.2.2. PD: discussion

II.1.2.3.3.2.2.1. Two normative systems to judge soundness and fallaciousness

We have seen that in the most recent version of PD, in which rhetorical concerns are addressed via the development of the notion of strategic manoeuvring, fallacies are interpreted as contextual infringements of context-independent rules of a critical discussion. This particular conception could turn out to be problematic in one respect: it has to do with this double system of norms we have at our disposal in order to judge the (non)fallaciousness of an argument. If "all derailments of strategic manoeuvring are fallacious and all fallacies can be regarded as derailments of strategic manoeuvring" (van Eemeren & Houtlosser 2002: 142), this means that all derailments of strategic manoeuvring are also fallacies in the pragma-dialectical sense, i.e. those argumentative moves which violate the rules of the critical discussion. Consequently, there is definitional correspondence and there should also be theoretical compatibility between the criteria by which we judge (dialectical) fallaciousness and those we use to judge derailments in strategic manoeuvring. In other words, any violation of the rules of the critical discussion should lead to a violation of the contextual conditions of sound strategic manoeuvring, and vice-versa. Yet these two sets of rules could be envisaged as incompatible: dialectical standards are taken in PD to be absolute, invariant, and compulsory for a proper critical discussion to take place; as for

norms of sound strategic manoeuvring, they are believed to be context-dependent and thus relative. The system, as it were, turns out to provide two epistemologically incompatible sets of criteria governing fallacy judgments.

On the one hand, the rules of the critical discussion state prohibited moves which, if realised, *ipso facto* render the argumentation fallacious; these are absolute rules on which the entire normativity of the system rests. On the other hand, we have a flexible context-dependent system of norms of fallaciousness which helps us determining whether strategic manoeuvring has derailed or not, the result of such derailment being a fallacious move. The question that emerges is therefore the following: how can we reconcile two sets of norms which seem to be contradictory in their very nature, but which nonetheless are designed to assess the same phenomenon (i.e. fallaciousness)?

One way of answering would be to weaken the idea that a derailment in strategic manoeuvring *necessarily* results in a violation of the rules of the critical discussion, and vice-versa, and thus to reassess the claim according to which fallaciousness and soundness are regulated by both systems *simultaneously*. This would allow both systems to co-exist without interfering with each other. Since norms of sound strategic manoeuvring are context-dependent and *ipso facto* make room for the possibility of *ad hoc* and conventional standards being adopted by particular communities of arguers, there is no principled objection to them adopting a rhetorical standard which happens to conflict with one or more than one of the 10 rules of the critical discussion. As way of illustration, we could imagine a community of arguers made of literalists who have decided to argue only on the basis of what is explicitly – or literally – said by conversational participants and that they would therefore not comply with rule # 5 which states that people are responsible for the unstated premises they leave implicit.[128] In such a case there would be a clash between the pragma-dialectical rules and the context-dependent norms of strategic manoeuvring; possible ways out of this paradoxical situation would include considering that a same argumentative move may be considered to be fallacious under pragma-dialectical standards but sound under strategic manoeuvring ones, or accepting that the on-going interaction, due to contextual parameters, cannot be reconstructed as a proper critical discussion. In the hypothetical example of the community of literalists, the agreement would have to be expressed in the opening stage and would result in a limitation of the application of dialectical rules, this itself resulting in a sub-type of critical discussion, as it were, where rule # 5 would not obtain for conventional reasons.

[128] This would run contrary to the predictions of Relevance Theory about natural communication, but the hypothetical example is precisely meant to stress that an argumentative practice can specify certain norms of argumentation.

F. Snoeck Henkemans[129] considers that whenever a situation like that occurs, contextual norms prevail over the pragma-dialectical rules of the critical discussion, but that in order not to be fallacious, the critical ideal of reasonableness should be kept intact, that is, that participants should still be willing to arrive at a resolution of the difference of opinion, albeit with specific conventional and *ad hoc* rules which may depart from the ten commandments. This means that, in the end, strategic manoeuvring that runs contrary to the ten commandments is acceptable as long as participants conventionally agree that the argumentative move in question is reasonable. Granted this situation, we have two options: either we accept that i) the pragma-dialectical rules of the critical discussion can be bent without this resulting in a fallacious move, but then they lose their universal nature and their observance can no longer constitute an acceptable criterion to ensure critical reasonableness, or ii) we stick to the pragma-dialectical rules but try to improve the model in another way so as to overcome the flaw. To my knowledge, PD has neither considered (i) nor (ii) at the moment. Yet, the interrogation yielded by the problematic example I just provided should be addressed, for failure to address it would amount for the whole system to predict that some norms are always the case, except when they are not. A recent volume (van Eemeren, Garssen & Meuffels 2009) presents experimental evidence in favour of the idea that people have intuitive judgments of reasonableness which seem to correspond to the pragma-dialectical rules of the critical discussion; the experimental design involved presenting people with several fallacious and non-fallacious arguments and asking them whether they thought the arguments were reasonable or not. The results show that people's judgements about these issues seem to confirm the ideal of critical reasonableness. Yet, this does not invalidate the concern mentioned above; rather, it indicates preference for the pragma-dialectical rules (which I would oppose to contextual and relative standards of argument validity). Perhaps additional research would need to assess how both sound moves and derailments in strategic manoeuvring, as predicted within PD, are perceived in terms of reasonableness, so as to be able to compare the judgments collected for both systems of norms.

In order to enlighten how both normative systems interact, an analogy with law is sometimes summoned by pragma-dialecticians.[130] The idea is as follows: just as discussions in the court of law are meant to provide an interpretation of certain deeds in terms of legal norms (i.e., a normative system) so as to evaluate whether the deeds of the accused indeed correspond to a criminal offence and what kind of sanction will be given to the accused should s/he be found guilty, the contextual norms of strategic manoeuvring will tell you whether the dialectical rules

[129] Personal communication.

[130] Marcin Lewinski, personal communication.

fully apply in the evaluation of the argumentative move in question. In other words, the verdict in the court of law has to confront legal principles to the facts and see whether the context in which the facts took place allow the facts to count as violations of these legal principles; by analogy, fallacy evaluation has to confront the rules of the critical discussion to the argumentative move under evaluation and see whether the contextual features and requirements of the activity type in which the argumentative move occurred allow the latter to count as a fallacy in the dialectical sense. In a way, the contextual norms associated to activity types are conceived as means to interpret whether the universal rules for a critical discussion apply or not.

The idea that contextual norms of activity types determine whether the rules for the critical discussion apply would actually be a way of specifying whether the full normativity of the critical discussion model has to be taken into account in a particular case, by allowing to suspend one or more rules without altogether compromising the ideal of reasonableness. The contextuality of fallacies therefore makes the description of argumentation much more prone to reflect actual argumentative practices, and this way, the possibility that contextual judgements may overrule the ten commandments is catered for.

While the analogy with legal discussion can shed light on the link between critical discussion rules and context-dependent peculiarities associated to activity types, there is still a crucial difference between both terms of the analogy. Legal discussions are subjected to conventional norms which have been agreed upon and recorded in constitutional documents, while argumentative practices in principle afford a cognitive account; as long as we consider argumentative practices to be the result of a social game played according to conventional rules, which is a conception PD is explicitly committed to, the question of normativity can be accommodated in the way outline above. Nonetheless, if we take argumentative practices to be a consequence of what happens in our minds in terms of information-processing, then it should be pointed out that cognitive principles are not comparable to conventional legal agreements because the former are not the outcome of deliberation – they are facts about human nature. Should we aim for a cognitive account of argumentation, the issues discussed here are actually not so relevant: while fallacy judgements are a matter of interpretation (argumentative moves have to be judged as acceptable or unacceptable relatively to a normative system), explanations of how argumentation is dealt with by our cognitive systems are not inherently related to the interpretation of conventional norms – they are as such likely to be adequate or inadequate explanations, regardless of a conventional normative system. The questions that I will address in Chapter III are precisely meant to shed light on the cognitive underpinnings of argument processing, and in this sense, they are unrelated to a normative discussion.

The question of the norm of argument validity is perhaps the one that has generated most debates in argumentation theory. It is not my purpose to take the discussion on this issue any further here; what seems to be noteworthy, though, is that the debate is still not settled among argumentation theorists. PD favours a double normative system; logicians favour formal logic criteria (such as deduction and non-contradiction) of validity; other informal logicians in the rhetorical vein privilege norms of argument acceptability that are audience-dependent. But virtually none of these approaches considers the cognitive aspect of argumentation, which is I think something that should be done, even more so if it is a naturalistic account of how our minds process argumentation we are after. This in turn will open up the path for a discussion of *cognitive* norms of argumentation.[131] Before I come back to this issue with a reflection on the relevance of the central question argumentation theories seek to answer, namely the question of knowing what makes arguments sound, valid, or acceptable, let me add a few considerations pertaining to my exposition of certain aspects of PD that could enrich the type of account of argumentation theory I am after.

II.1.2.3.3.2.2.2. Pragmatics in PD

Pragmatic aspects of PD can further be commented on should we want to adopt a naturalistic perspective on argumentation. In particular, since PD relies on early pragmatic theories of communication it is *ipso facto* concerned by the recent evolution of research in contemporary Pragmatics.[132]

The first point of discussion has to do with the account of meaning provided in those theories. Before one can process an argument in terms of its accessibility, one needs to understand it; in other words, comprehension is a necessary condition for argument evaluation. This Aristotle pointed at indirectly by isolating the category of fallacies depending on the problematic use of language (ambiguity, amphiboly, combination, division of words, accent, form of expression), which were to be named afterwards the *in dictione* fallacies. Making sense of a sequence of discourse is required prior to any validation procedure, for one cannot reasonably evaluate something which one has not understood. So, intuitively, a full-fledged theory of argumentation

[131] Here the term 'norm' should not be taken as an ideal meant to be followed by arguers, but rather as a quasi-biological, hard-wired standard governing the way we process communicated information. I intend to explore an alternative direction which addresses the question of knowing how, somewhat by default, arguments are processed by our minds, and contend that the mechanisms governing argument processing are no different from the general mechanisms of information processing.

[132] See Oswald (2007a) and (2007b) for an extended discussion of these issues, on which this section is based.

should address the question of *interpretation*, or at least include some module on which it can rely to deal with the question of meaning construction. PD tackles the issue by resorting both to SAT and to an elaborate version of Grice's framework.

SAT holds that an addressee is able to fully understand a speech act if he is able to grasp its illocutionary force, i.e., when he is able to know what kind of speech act is at stake (assertion, promise, request, order, and so forth). In order to identify the speech act that has just been performed by the speaker – and thus to understand it – he will need to identify the category the speech act belongs to. To do this, a hearer must recognise the speaker's *intention* (in the Gricean sense); the means by which he does so are assumed to be conventional, i.e., it is because we know by convention that specific verbal expressions are used to achieve specific effects (such as understanding which speech act has just been performed) that we are able to recognise the speech act. According to Searle, communication will be successful "if the hearer understands the sentence, i.e., knows its meaning, i.e., knows the rules governing its elements" (Searle 1969: 48), these rules being conventional. Sperber and Wilson (1995) raised a few concerns about this take on communication, and since PD elaborates on SAT in order to define argumentation as a complex speech act (also referred to as a "speech act complex" in Chapter 2 of van Eemeren and Grootendorst 1984), this debate is also of importance to PD.

First of all, one and the same sentence can correspond to different speech acts depending on the context in which the speech act is formulated. Nevertheless, the answer provided by SAT does not go into detail to explain how and why this can be the case. Take for instance (60):

(60) We wouldn't want this subject to be mentioned in Laszlo's presence.

Only contextual cues will tell us whether (60) is to be understood as request, an assertion, an order, an advice or even a threat. In cases where no linguistic or prosodic features favour one interpretation over another, pragmatic approaches will usually say that contextual information takes over and helps us inferring the right meaning. In particular, a Searlean account would assume that such is the case for (60), provided the "utterance in a context can indicate the satisfaction of an essential condition without the use of the explicit illocutionary force-indicating device for that essential condition" (Searle 1969: 68). This tells us what happens (namely, that the context can make for the lack of explicit illocutionary force markers and allow to satisfy the speech act's essential conditions), but does not provide concrete details about the procedure followed by the hearer for this to happen. In order to overcome this weakness, Searle drew on the Gricean framework to address the question of indirect speech acts.

Sperber and Wilson (1995) discussed both Speech Act Theory and the Gricean approach. Their arguments regarding SAT come down to the fact that it lacks explanatory adequacy:

> "It is one thing to invent, for one's own theoretical purposes, a set of categories to use in classifying the utterances of native speakers, or to try to discover the set of categories that native speakers use in classifying their own utterances. It is quite another to claim that such a classification plays a necessary role in communication and comprehension." (Sperber and Wilson 1995: 244)

They point out that there are no clear reasons to claim that the fact that the hearer is able to place a speech act into a category plays a decisive role in comprehension: whether we identify a particular speech act is not a necessary condition for understanding it. Stating that we are able to classify a speech act as an advice does not say much as to why or how, but mainly whether we are able to do it. Sperber and Wilson take the example of a tennis player and make the following comparison: recognising a lob, a volley, a backhand or a smash is not a sufficient condition to perform one, let alone to adopt adequate behaviour to deal with it. From a cognitive point of view, speech act classification and identification by the hearer may well constitute an additional and unnecessary layer of information. This is the reason why Relevance Theory argues that calling upon conventional reasons and establishing a classification without explaining how we use it does not, from a cognitive perspective, shed enough light on the crucial question of the nature and role of context construction in mechanisms of interpretation.

Sperber and Wilson (1995) also discuss the foundational works of H.P. Grice, and in particular his model of interpretation. The strong assumption of this paradigm is the idea that hearers are able to grasp explicit and implicit meaning through some kind of *calculation*. The hearer does this by following a rational step-by-step procedure known as the *working-out schema*.[133] Wilson (2000, 2003) recalls however that this explanation of meaning construction faces serious

---

[133] Grice's "working out schema" for conversational implicatures:

> (a) He has said that p.
>
> (b) There is no reason to suppose that he is not observing the maxims, or at least the CP [= Co-operative Principle].
>
> (c) He could not be doing this unless he thought that q.
>
> (d) He knows (and knows that I know that he knows) that I can see that the supposition that he thinks that q is required.
>
> (e) He has done nothing to stop me thinking that q.
>
> (f) He intends me to think, or is at least willing to allow me to think, that q.
>
> (g) And so he has implicated that q.

(Taken from Wilson (2000: 416). Also found in van Eemeren and Grootendorst (1984: 120)).

problems of cognitive plausibility. Notably, according to Wilson, it seems unlikely that small children, who seem pretty good with implicature derivation, go through such complex procedures.[134] Moreover, it fails to explain exactly how an implicature is retrieved – the conversational maxims function as norms according to which implicatures are derived, but these norms are not *justified* in any way – and thus merely shows how "once constructed, it [the implicature] might be confirmed as part of the speaker's meaning" (Wilson 2000: 419). As a matter of fact, it is partly in reaction to the issues faced by SAT and by Grice's framework that RT's model was built.

In addition to these *upstream* issues which have to do with the cognitive operations that take place prior to argument evaluation, there are also *downstream* issues which refer to the output of argument evaluation, namely issues of conviction. Pretty much as most approaches to argumentation, the standard Pragma-dialectical model implicitly assumed that sound arguments should be convincing and *ipso facto* lead to belief fixation; yet, this called for further elaboration, which has partly been supplied by the extended model of strategic manoeuvring through a reflection on rhetorical issues.

Common sense would make us expect sound arguments to convince (leading *ipso facto* to belief fixation) and fallacies to fail to do so; and indeed, this is a scenario that can be observed in naturally-occurring arguments. But things are more complicated. As a matter of fact, experience shows us that sound arguments may fail to convince, while fallacies may succeed. Consider the following examples:

(61) a. Winston: "Why are you washing the potatoes if you are going to peel them anyway?"

b. Laszlo: "Let me do things my way".

(62) Four million Japanese people cannot be wrong. That was the number of people that rushed to purchase *Dragon Quest* on its release in Japan. (BBC)[135]

Reconstructing (61), it appears that Winston's standpoint is that Laszlo should not wash the potatoes. This statement is sustained by the argument that dirt comes off the potatoes when peeling them, and therefore follows the conclusion that it is unnecessary to wash the potatoes

[134] See Wilson (2000 and 2003) for a discussion based on evidence from psychological experimentation involving children. Wilson quotes, among others, Bretherton (1991), Newcombe and Zaslow (1981), Tomasello, Farrar and Dines (1983), O'Neill (1996).

[135] Taken from "Dragon Quest fires up gamers", article by Seth Goolnik, April 14, 2006. Available on: http://news.bbc.co.uk/1/hi/technology/4907668.stm.

before peeling them. At first glance, this argument is sound, and could be pretty convincing. But what if Laszlo wants to deal with potatoes the way he was taught to by his parents (i.e., washing them first)? What if he considers that it is more pleasant to peel clean potatoes than potatoes covered with dirt? There is a possibility that these alternative options could overtake sound argumentation and lead Laszlo to reject the representation yielded by Winston's statement, even though the supporting argument was valid. More generally, it is quite obvious that belief fixation does not necessarily follow from valid argumentation – simply because many other parameters enter belief fixation/rejection processes – and therefore that there are important nuances to add to the intuitive idea that sound argumentation makes its conclusion convincing.

Conversely, (62) may well not be ruled out as non-convincing, though it exploits the fallacy known as the *bandwagon fallacy*, or *argumentum ad populum*. The context is the release of an awaited videogame in Japan. By uttering (62), the journalist is actually communicating positive attitudes as to the game's quality. I can easily imagine that despite being fallacious, this type of argument can weigh in someone's decision to buy the game ("If 4 million people bought it, it has to be good, thus it's worth buying it too"). Intuitively at least, it could very well be the case that the mind actually manages these kinds of fallacious arguments as acceptable informative articulations, i.e., as somehow valid in the given context. This coincides with the idea that the mind does not always follow critical and logical pathways when coming to entertain a belief as true; psychological and sociological studies, such as Milgram's famous experiments on obedience (Milgram 2004 [1974]), or Festinger's work on the notion of *cognitive dissonance* (Festinger 1985 [1957]) have indeed showed that other parameters can influence people's rational beliefs and behaviour.[136]

This is not to say that PD does not *explain* things; in fact PD has both *micro* and *macro* distinct levels of explanation. *Micro* explanations rest on the norms governing speech act performance and recognition. They are internal in the sense that they are supposed to reflect what language users do when communicating, but they are not internal in a cognitive sense, since they do not deal with cognitive mechanisms of interpretation. *Macro* explanations (i.e., the rules of the critical discussion) are external to language, since they reflect the way the theorist believes the activity of arguing is regulated according to certain philosophical standards. In this sense, the model specifies what argumentation should be. Of course, one could claim that knowing the rules that determine sound argumentation should allow us to perform argumentation properly, and indeed, if we follow the rules of the critical discussion, what we get is sound argumentation. But establishing a set of rules without reliance on cognitive evidence, as supplied today by

[136] See §II.2.1.1. and §II.2.2.2. below.

numerous studies within the field of experimental pragmatics and cognitive psychology (see for instance Noveck & Sperber 2004, Pohl 2004), does not guarantee that these rules reflect what our mind does.

The main difference between PD and cognitive pragmatics I evoked, in terms of their analysis of how language users process information, brings forward one of the principles from which argumentation scholars may benefit: addressing cognitive aspects of argument processing should entail addressing dynamic aspects of *online* 'as-we-speak' interaction. In addition to dealing with properties of arguments and properties which are inherent to the procedure of arguing, such approaches should allow for predictions on real-time argumentation to be made, which would be based on *internal* insights. What I mean by this is that arguments, like any other ostensive-inferential stimulus, mobilises our cognitive abilities, and that insight into these cognitive abilities should provide an extremely valuable input in theories of argumentation. This will precisely by one of the last points of discussion in Chapter III.

II.1.2.4. Limitations of formal and informal logic for a naturalistic account of fallaciousness

II.1.2.4.1. Objections to formal logic

According to the literature on informal fallacies, to which studies in informal logic devote significant attention (e.g. Walton 2008, Johnson 1996, Johnson & Blair 2000), some argumentative flaws which are characterised as informal fallacies are not to be classified as such from a logical point of view. Informal fallacies have been dealt with as those errors of reasoning which do not result from a formal fault;[137] by definition the success of informal fallacies does not build on formal mistakes of reasoning but rather on a variety of constraints operating elsewhere than specifically on the formal articulation between premises and conclusion. This is for instance the case of all the so-called 'red herring' fallacies that operate by getting the conversational participants to lose track of the standpoint that is being called into question through the argumentation. Sometimes this goal is attained by changing topics: an effective *ad misericordiam* fallacy (or appeal to pity) can make the conversation derail onto a sub-discussion about the emotional issues brought forward by the content of the argument. Informal fallacies can also force participants to digress or to defend themselves: the *ad hominem* fallacy, for instance, through an attack launched on the arguer, may compel her/him to defend her/himself instead of sticking to the argument and as a consequence it might misdirect the initial discussion onto

[137] See Whately (1843: 155) for a post-Aristotelian classification of fallacies; he draws a clear distinction between logical and non-logical (or material) fallacies. The latter have traditionally been referred to as 'informal fallacies' in subsequent literature on the topic.

another topic, leaving the initial argument unattended. A departure from the assessment of the formal properties of an argument may also be brought about by making the participants unwittingly accept unwarranted starting points of argumentation (e.g. the *straw man* fallacy, which is about getting a misrepresentation of someone's standpoint to be accepted as the actual standpoint s/he formulated in the discussion).

These are typical examples of informal fallacies showing that there are some argumentative moves which are detrimental to the process of arguing properly and that these cannot be handled by the tools of formal logic, since they operate in the periphery of the argument's formal articulation, thus falling out of the scope of aspects covered by formal logic. Yet, such issues need to be addressed for a theory of argumentation – and of manipulative communication as well – to be comprehensive, and this would arguably be achieved only by taking the explanation beyond the limits of formal validity.

The example of the fallacy known as *begging the question* is particularly well-suited to illustrate this last point. I have mentioned (see footnote 111 above) that to a certain extent, this particular type of fallacy can be considered as a valid argument, albeit a trivial one, as it involves a tautology. From a logical standpoint this is acceptable, since every time the premise is true, so will the conclusion (given that both are the same); yet, from the point of view of the pragmatic functions of argumentation this type of argument cannot be accepted, because i) one cannot reasonably be taken to having proved anything by assuming the proposition to be proved as part of the premises brought forward to sustain the very same proposition, and ii) there is no clear informative benefit – and thus no element of justification – gained from repeating the same proposition in case we intend the repetition to act as a justification. Accordingly, what we need to overcome a logical account of arguments of this type is an account designed to explain why exactly we intuitively feel that tautologies do not constitute acceptable arguments.

A second argument pointing to a limitation of formal logic (keeping in mind that what we are interested in are the conditions favouring the elicitation of consent) is that the argumentative nature of certain exchanges is sometimes left out of the literal form, and this is something formal logic has trouble dealing with, as it cannot capture the argumentative link existing between two clauses when this link is only pragmatically – or contextually – expressed. In this respect, let me recall an example provided by van Eemeren & Grootendorst (1992):

(63) Let's take an umbrella, or did you want to get wet? (van Eemeren & Grootendorst 1992: 47)

In (63), it is possible to imagine a context where the question following the proposal has to be interpreted as an argument meant to sustain the claim that the speaker and the hearer should

take an umbrella. However, no explicit argumentative pointer favours this interpretation (the operator 'or' does not semantically encode causality, but disjunction). Van Eemeren and Grootendorst explain the causal relationship by arguing that the first clause, although it is not literally a standpoint (you cannot answer "I disagree" to "let's take an umbrella"), functions as one, and that the second is not just a yes/no question. I would add that the alternative offered in the second clause is somewhat ridiculous (most people would not at first glance be inclined to get wet); as such, it is easily disposable, and reinforces the initial proposal. Interestingly enough, we cannot even decide whether (63) carries argumentative force or not from a strictly semantic point of view. What if the speaker considered that the hearer had reasons to get wet (because s/he likes it, or for any other possible reason)? In this case, (63) would not be argumentative, but strictly interrogative. The point is that we cannot explain why (63) can be argumentative by virtue of its semantic or logical properties without calling upon the context. And thus it is only *pragmatically* that we can infer the argumentative nature and function of the utterance, and examples such as (63) lead us to the conclusion that a comprehensive account of argumentation cannot be handled by formal logic alone, as the latter will typically not take into account pragmatic aspects of communication.

A third argument favouring a departure from an exclusive formal logic account of argumentation relates to the difference that some researchers have established between formal and practical reasoning, which basically grew out of the observation that real life argumentation seldom conforms to formal standards of validity. This observation is also traditionally associated to the motivations upon which the advent of informal logic has built, notably through Toulmin's (1958) *Uses of argument*, and Perelman & Olbrechts-Tyteca's (1958) *La nouvelle rhétorique*. The two accounts grew dissatisfied with formal logic, insofar as they noticed that language users did not necessarily always comply with rational standards of formal logic and that people very often seemed to be convinced by fallacious arguments. Taken seriously, the observation is potentially devastating for formal logic accounts, should the latter have any psychological pretensions: if logic is assumed to govern the way we rationally argue, then we must find an explanation for empirical data showing that these standards remain unobserved in many situations.

I have mentioned earlier that many errors of reasoning which are labelled in logic textbooks as fallacies but which do not build on formal errors do influence the course of the argumentative exchange by distracting the participants from the actual argument at stake. Let us elaborate on this in order to illustrate the point made in the preceding paragraph. For instance, if a speaker is asked to justify why she took such and such decision, but instead of answering she files an *ad hominem* attack on the questioner, the course of the conversation runs the risk of being diverted into a sub-discussion about the attack, as the questioner might feel obliged to defend herself.

Mechanisms of this kind in principle have nothing to do with the form of the argument, as they are precisely meant to take participants' attention away from it. In fact, one could even potentially reconstruct the structure of an *ad hominem* argument so as to make it adopt a syllogistic form, thus ensuring that any concern about form is dismissed. Imagine for instance that I tell my friend Laszlo something like (64). The argument I am putting forward could be reconstructed as stated in (65):

(64) I won't trust you with my money because of your past misdemeanours.

(65) a. People who have been condemned in the past for misdemeanours should not be trusted with money.

b. You have been condemned for misdemeanours in the past.

c. You should not be trusted with money.

The sequence in (65) is formally valid as it takes the form of a *modus ponens*: if both premises are true, the conclusion thus necessarily follows. Yet, there are grounds to question the truth of the first premise (65a); we only need one counter-example to prove it is false and by the looks of it, it indeed seems to be a rather easily falsifiable assertion. But this is not the main point of the discussion of this example; when used in a conversation, this type of fallacy – in addition to attacking the interlocutor through an unflattering description of his behaviour or personality – may play an alternative role, that of prompting the target of the personal attack to feel compelled to respond, hence risking to lose track of the standpoint that was called into question (i.e. (64), "I won't trust you with my money"), which was the issue originally under discussion. The significance of the fallacy in such cases is not so much the debatability of its premises than its ability to generate a departure from the point originally under discussion. This is especially vividly illustrated by many examples of the *tu quoque* variant of the *ad hominem* fallacy, in which the speaker "undermines the other party's credibility by pointing out a contradiction in that party's words or deeds, for example a contradiction between their opinions in the past and in the present, or between what they say and what they do" (van Eemeren, Grootendorst & Snoeck Henkemans 2002: 112).

Through a *tu quoque*, the speaker reverses the attack and tries to get her opponent to defend himself by highlighting one of his own inconsistencies.[138] This in turn might induce a defensive move relating to the new attack, which leaves unattended the issue that was at first under discussion, this potentially constituting one of the goals the speaker who launched the *tu quoque*

[138] Think of those discussions in couples, where a reproach is answered by another reproach, without contributing to the resolution of the conflict of opinion at stake.

was seeking to attain. Example (65) shows that an *ad hominem* argument may be formally valid; yet, the fallaciousness of such an argument is to be sought on a different level; it is also both by questioning the acceptability of premises and by assessing what type of response an argument may induce in terms of conversational moves that a thorough account of fallacious reasoning can unfold. Such parameters are important when it comes to influencing the consent one might give to a proposition, or when it comes to getting the addressee to shift his attention onto something else – even more so when this is done in a covert way, i.e., when the speaker does not communicate the actual intention motivating her argumentative move.

It is thus on these grounds that informal logic developed, primarily fed by the ambition of describing and explaining an argumentative reality which seemed to exceed the scope of formal logic's capabilities, and it is also for this reason that a shift from *rational* to *practical* reasoning was observed in argumentation studies. The issues of consent and the parameters that might influence are not captured by formal logic. The reason for this is fairly intuitive: a superficial observation of real-life argumentative situations reveals that people are convinced through different means, and not only by relying on formal norms of validity; one can believe a proposition because someone they trust as an authority tells them to (as illustrated by the famous Milgram experiment, cf. Milgram 1974)[139], which is a mechanism that may bypass operations of formal validity checking. This is in fact the very principle of the *ad verecundiam* fallacy in which the authority of an expert is called in to support and legitimise the acceptability of a claim. Conversely, a hearer may decide not to believe what the speaker says, not because the reasoning involved is formally invalid, but because he does not trust the speaker. This effect can be observed in occurrences of the *ad hominem* fallacy, which does the exact opposite of the *ad verecundiam*, as it is meant to lower and if possible destroy the credibility of any source (including the speaker) whose authority is used to legitimise a claim. All these issues, which definitely deserve proper treatment within argumentation theory, are among the central concerns of informal logic as they give us insights into how people process argumentation; accordingly, the import of informal logic is of great interest for the topic of manipulative communication, to the extent that fallacies may be deliberately produced in order to elicit a consent that would probably be questioned should the argument be critically evaluated.[140]

To round up the arguments against logic brought forth by informal approaches to reasoning and argumentation, let me recall what I mentioned earlier, namely that argument validity was

[139] For an exposition see section §II.2.1.1.

[140] I will develop my own vision of the relationship between fallacies and manipulation further along (see §III.5.1.).

probably the central question of argumentation studies. In light of what has just been discussed, a few precisions should be adduced. First of all, behind the question of the norms according to which argument validity – or acceptability – should be judged lies a crucial question: why do people believe propositions and deem certain argumentative chains valid? Or, better put, what are the parameters that play a role in processes of conviction/belief fixation in communication? If in principle there is no strong reason to doubt that formal validity is a reliable cue to hold as true the conclusion of an argument, this does not mean that there are no alternative sufficient conditions, other than formal validity, to convince someone of something, to the extent that one may believe what the speaker is asserting just because one trusts the speaker's judgement and credibility, because all the members of the community believe the same thing or because intuition tells us that the standpoint is correct, without even engaging in a verification procedure meant to critically test the standpoint, and so on. Second (and consequently), one of the central issues in argumentation theories is the investigation of the conditions, besides formal validity, that play a role in the obtainment of consent. As Dascal observed, "our 'natural reasoning' often deviates from the norms of correct reasoning". We should thus be out for an account of "a wide range of ways of extending our knowledge that cannot be handled by formal logic alone" (Dascal 2005: 5).

One of the areas of research where logical norms of validity have been tested is cognitive psychology. The next section will review some of the arguments that have been put forward in support of the claim that formal logic perhaps does not map our rationality in the best of ways, thus pointing to a limitation undermining its use in the construction of normative criteria of argument validity.

II.1.2.4.2. Cognitive issues

One of the most important scientific issues concerning formal – but also informal – logic is that of the cognitive plausibility of their models. As both inquire about the conditions defining norms of correct reasoning, it should be expected that the accounts they develop provide accurate models of how people reason. Specifically, the models provided by these disciplines should in principle be representative of the argumentative mechanisms at play in our minds.

A few reservations can be formulated with respect to this claim through an assessment of whether formal and informal logic manage to achieve this goal. Let me first start by reviewing some arguments against formal logic.

The goal of formal logic is to provide means of guaranteeing formal validity; accordingly the system it develops can be said to advocate a failsafe procedure meant to make sure that invalid sequences of propositions are left out of the model. In other words, formal logic tells you which

conclusions are licensed, and which are not of by virtue of their formal invalidity. Characteristically, formal logic is systematic and yields consistent outputs, insofar as the procedures involved in deductive reasoning are always the same and always yield the same result, regardless of the variables, and make no room for contradictions. It is probably because of their rigour that logical models are the first that have been implemented in computers (see Huth & Ryan 2004): the calculations they involve are in principle a guarantee of flawless reasoning.

Yet, arguing for the perfection of flawless reasoning is quite different from defending the idea that these logical procedures actually do map the way our mind functions when processing arguments; this last idea faces serious challenge. The first argument against this view lies in the fact that we humans are prone to make reasoning errors, and, what is more troublesome, that we may fail to notice that these errors ought in principle to make us revise our reasoning and dismiss its output as unreliable. Many trends in psychological research, over the last 40 years, have studied the errors the human mind is likely to make in specific tasks. One of these paradigms, whose development has been initiated by psychologists Daniel Kahneman and Amos Tversky, focused particularly on judgemental errors people make when making decisions, in an attempt to shed more light on the theory of rational decision.[141] In particular, they showed in a series of publications that people happen to rely on certain types of *heuristics* when asked to provide probability judgments, among which the representativeness and the availability heuristics (cf. Tversky & Kahneman 1974). To take an example, they showed that when asked questions of the type "What is the probability that object A belongs to class B?" and "What is the probability that event A originates from process B?" (Tversky & Kahneman 1974: 1124), subjects typically answer based on the degree to which they perceive a resemblance between A and B (representativeness heuristics), which is not sufficient evidence to conclude that a causal relation obtains between A and B. Their idea is roughly that the mind does not function like a computer which systematically and consistently processes input information, but that it takes shortcuts, and often yields outputs through fallible procedures which involve risk and thus may lead to judgemental errors.[142]

I will not expose the details of Kahneman and Tversky's work here, but let me note that this body of research constitutes a strong empirical basis to support the intuitive and common idea that we humans are fallible information processors, and that we may sometimes – due to

[141] See e.g. Kahneman & Tversky (1972, 1973) and Tversky & Kahneman (1974, 1983).

[142] For a detailed discussion of judgemental errors, see Kahneman and Tversky's seminal paper (1974), and the eponymous book (1982) by the same authors.

economic constraints of efficiency among other parameters – miss some information, or at least neither look for nor take into account all relevant information when processing inputs. These observations will be of capital importance for the rationale of the model I will propose in Chapter III.

Although Kahneman & Tversky's research was aimed at contributing to a theory of rational decision, there is *a priori* no reason to consider that such heuristics may not also apply to other rational processes, such as language processing. This idea is taken up by Nicholas Allott within Relevance Theory as he argues that "reasoning involves shortcuts, many of them heuristic […]: non-algorithmic procedures which do not guarantee reaching the right answer" (Allott 2007: 252).[143] In particular, his aim is to show that the processing of utterances may take these 'shortcut' paths, sometimes to the expense of consistency and logical validity, under a global constraint of efficiency.[144] Sperber, Cara & Girotto have earlier hypothesised something along these lines, when they argued that "rationality as generally understood involves logicality but has other components too, for instance the ability to allocate one's cognitive resources efficiently, or that of performing non-demonstrative inferences" (Sperber, Cara & Girotto 1995: 44). The idea summarised in this last sentence thus corresponds to the first argument one may use to object to the use of formal logic as an exclusive model of our natural reasoning mechanisms.

A series of experiments first carried out in the 1960s by cognitive psychologist Peter Wason is also worth mentioning here, as it specifically focused on people's ability to perform logical reasoning of the *modus tollens* and the *modus ponens* type. The original experiment, dubbed the "selection task" is, according to some, "the most intensively researched single problem in the history of the psychology of reasoning" (Evans, Newstead & Byrne 1993: 99), and was originally designed to test a subject's performance in logical tasks of reasoning.

The experiment goes as follows: a subject is presented with 4 cards and is told that each of these cards has a letter on one side and a number on its back. Two cards in front of the subject are labelled with a letter (say A and G), and the remaining two with a number (say 7 and 8). Subjects are then asked to answer a question formulated roughly as follows: "indicate which of these cards you need to turn over in order to judge whether the following rule is true: if there is an A

---

[143] Saussure (personal communication) notes in this respect that there is a contradiction in terms in the expression 'non-algorithmic procedures', in that procedures, in the interpretative sense, are necessarily algorithmic. Allott's quote should thus be revised in this respect, although the idea underlying still applies, to the extent that he intended to point out the fallibility of human reasoning.

[144] This is one of the underlying ideas of Relevance Theory (Sperber & Wilson 1995).

on one side, there is a 7 on the other side". It is assumed that reaching the correct answer requires logical reasoning, as the rule is formulated as a classical *modus ponens* of the form 'P → Q' ('If there is an A on one side of the card, then there is a 7 on the other').

The proper way to go about this task is to reflect upon the situations in which the rule would be violated, i.e., subjects should look for scenarios in which P is the case and Q is not. Only choosing scenarios confirming the rule would not constitute definitive evidence that the rule is correct, since there is in principle always a possibility that one of the cards we did not turn actually disproves the rule.[145] The correct solution to the Wason selection task is thus that the A card and the 8 card should be turned, since those moves are the only ones that will tell us if the disproving scenarios evoked above obtain. If we turn card A and we find a number other than 7, then this will be direct evidence that the rule is not verified. Most subjects get this one right. Now, if we turn the 8 card and find an A on the other side, this will also count as one of the scenarios in which P is the case and Q is not, and thus turning the 8 card is also instrumental to assessing the validity of the rule. But the majority of tested subjects do not get this one. Instead, they go for the card labelled with a 7; turning this card, in case we find something else than an A, will not disprove the rule; in fact, it will say nothing about the rule's validity, because all the rule says is that if (but not *only if*) there is an A on one side, you should find a 7 on the other, not that on the back of a 7 card you should find an A. The rule predicates that A cards will have a 7 on their back, but does not predicate anything of cards labelled with a 7; in fact, whether a card labelled with a 7 has an A or any other letter on its back is irrelevant to the rule. In abstract terms, the errors subjects make is to mistake a conditional statement for a bi-conditional one. Reasoning in syllogistic terms, turning the 7 card would amount to affirming the consequent, which is not an acceptable logical move to infer that the antecedent is the case.[146] Finally, nothing can be said by turning card G, as the scope of the rule excludes it (the rule is only about A cards), and indeed no one picks that one.

Results of this experiment show that people perform poorly in 'abstract' settings where they have to reason on operations which are not grounded in plausible real-life situations (no one would *a priori*, outside an experimental setting, come up to me and ask me to select cards in

---

[145] More generally this train of thought is typical of falsificationists (cf. Popper 1959, 1963), who consider that one can never prove that an explanatory proposition for a given phenomenon is absolutely valid since potentially there might always be an occurrence of the phenomenon that will disprove the explanation); the only way we shall make scientific progress is by providing refutations of propositions, as these constitute eliminations of possible explanations, thus reducing the set of potential explanations for a given phenomenon and getting us *ipso facto* nearer to the truth.

[146] Affirming the consequent is one of the *extra dictionem* fallacy identified by Aristotle.

order to verify whether a given rule is observed or not). Up to 80% of people do not come up with the correct logical answer. It was suspected shortly afterwards that the abstract nature of the task was precisely what was preventing subjects from performing satisfactorily, and so additional versions have been set up and tested, among which a deontic one, labelled as such because it deals with the application of a rule in a plausible 'real-life' scenario. The narrative describing the task was something like this: you are a police officer, and you go into a bar to make sure that the rule according to which no people under 21 can drink alcohol is being observed; at the bar, you see on person drinking soda, one drinking beer (but you don't know their age), one who is 18 and one who is 45 (but you cannot see what the latter two are drinking). And the question you have to answer is: who, among those 4 people, do you need to interrogate in order to verify that the law is being observed? In this setting, the vast majority of subjects perform as 'logically' expected (they select the person with the beer to check their age, and the person who is 18 to check whether they're drinking alcohol). Those two people actually correspond to the A and 8 cards in the abstract design, but in this deontic version, subjects have no trouble in performing the selection task correctly.

Sperber, Cara & Girotto (1995) have convincingly argued for a cognitively grounded explanation that not only explains past results in the Wason selection task, but also predicts further results in specific variants of the task. Their idea is quite simple: before they can perform the task they are asked to carry out, subjects need first to understand what is expected of them; and depending on how and where they find relevance in the instructions (which also depends on both the wording and the context of the experiment), their performance will vary along the logical/biased reasoning continuum. Sperber *et al*. construe the selection task not as a task of conditional reasoning requiring a logical competence, but a task of selection of relevant evidence to answer a question, which means that "the processes involved [in performing the selection task] are […] the standard processes of verbal comprehension" (Sperber *et al*. 1995: 36). When subjects fail, it is because they are over-confident, or "nearly-incorrigible pragmatic 'optimists'" (Sperber *et al*. 1995: 90) in the output of the pragmatic processing they went through to understand what was expected from them (i.e., they are wrong in determining their intuition of relevance). When they succeed, it is because the pragmatic meaning of the version of the task they are performing typically elicits logically-compatible intuitions (and thus yields logically correct selections).

In a series of four experiments, Sperber *et al*. (1995) showed that performance could indeed be improved in the selection task by manipulating the data presented in the rule and the instructions. The idea guiding these experimental designs was quite simple: all was needed to elicit good performance was to raise the expectations of relevance of the expected answer. In the

original 1960's experiments, instead of going for the 'P & ¬ Q' logical answer, subjects typically were going for the 'P & Q' answer. What Sperber *et al* did was to create contexts and formulate rules so as to strengthen the relevance of going for the expected – logical – answer. And they found that subjects are indeed able to perform well when such conditions are met.[147]

This goes to show that more than about logical reasoning, the Wason selection task is perhaps about processes of verbal comprehension. Nevertheless, even if the value of the experiment as a test for logical reasoning is consequently undermined, it still shows that logical reasoning is perhaps not a standard of rationality that people follow by default, for otherwise they would get it right in the first place.

Evans (2004) has summarised different explanations that have been advanced to account for errors in the Wason selection task by bringing into light the existence of different biases in reasoning that could be parasitic to the performance of logical reasoning, such as the confirmation bias and the matching bias. Beyond the explanatory adequacy of each of these accounts, this still has implications for our conception of logical reasoning and rationality:

> "there is now a big debate about logic and rationality. This debate complicates the story of biases in deductive reasoning somewhat. The reason is that 'biases' in this field of research are measured relative to logic. A bias is defined as a systematic not random error. It involves people either neglecting some logically relevant information, or being influenced by factors that have nothing to do with the logic of the task. Hence, a bias is only evidence of irrationality if you think that logic is the right way to measure rationality." (Evans 2004: 129)

If it is the nature of rationality we are concerned with, then there are two options: either being logical is the standard of rationality, but in that case most of us are irrational, judging by the widespread inability to perform the selection task successfully that has been identified by Wason and all his followers through numerous replications, or reasoning logically is not a standard of rationality, in which case we have to come to terms with the idea that formal logic is under-equipped to model the way our brains spontaneously and naturally perform reasoning operations. The results of the Wason selection task show that logic is perhaps not to be taken as a default reasoning engine; yet, it must also be acknowledged that once experimenters explain the correct solution to subjects who failed in the task, the latter are perfectly able to understand the logical calculation behind the correct selection of cards. This means that even if we do not

[147] It is noteworthy that the recipe they followed to design 'easily' solvable selection tasks tries to exploit the same principle that makes the deontic version of the task easier, namely boosting the relevance of picking the 'P & ¬Q' option (but in Sperber *et al*.'s experiments there was no deontic component).

engage in logical reasoning by default, we are well capable of doing so; accordingly, this is evidence that there is some psychological reality to formal logic, this not being equivalent to saying that logic is the only system of reasoning.

So where does that leave us? Recalling Dascal's quote above (§II.1.2.4.1.), it appears that even though we are capable of thinking logically, this is not necessarily what we do by default. A number of biases have been proven to exist (cf. Kahneman & Tversky 1974, Pohl 2004), and moreover they have been shown to interfere with logical reasoning. This should be taken as evidence that the standards elaborated by formal logic accounts of reasoning should not be taken for granted as accurate models of spontaneous reasoning. Of course, we are capable of reasoning logically, and what is more, the outputs of such reasoning are reliable; but this is not to say that we engage in that specific type of information-processing by default. In particular, when processing complex arguments – as informal logic showed –, we are sometimes prone to fork off the logical path; that, however, does not mean that we are being irrational. Rather, it is a clear indication that the way we reason, which is intimately coupled with the way we process information, bears additional constraints to that of logical consistency. It is precisely those aspects of information-processing in argumentation that I am interested in, as it is my conviction that they are responsible for the success of a manipulative attempt framed as (fallacious) argumentation.

II.1.2.5. Conclusion

Despite the tremendous amount of research carried out by informal logic approaches to argumentation, little cognitive work has been undertaken in the field; in this respect, the following quote by van Eemeren and Grootendorst is highly revealing:

> "There is no need to have detailed knowledge of all the cognitive processes that play a role in the interpretation of a discourse or text in order to be able to carry out an analysis based on externalized textual characteristics, but some insight into these processes can, of course, deepen the analysis." (van Eemeren and Grootendorst 2004: 74)

The purpose of the present study is precisely an inquiry about cognitive mechanisms of information-processing in order to identify conditions under which people end up believing things they would not believe in normal circumstances or under attentive critical scrutiny. I have discussed the main tenets of informal logic and concluded that the major objection these can be charged with is that of overlooking the fact that not everything that is relevant to argument evaluation is in the text, in its structure or in the context: it is my conviction that phenomena such as fallacies, which are realised in discourse, may work in deceptive enterprises precisely

because they have a cognitive reality; that is, fallacies do work because the way our cognition works allows them to work. This is an idea I will elaborate on in chapter III.

While both informal and formal logic theories provide extremely rich insights into the type of reasoning and the type of argumentative structures that can be involved in argumentation, they both fall short of answering one of the most important questions related to the study of argumentation, namely the question of knowing how exactly people produce, interpret and evaluate arguments in ordinary communication, that is, without either reducing argumentation to the notion of valid inference or broadening its scope so as to include all potentially admissible moves. In particular, the fundamental psychological question remains open: I have reviewed arguments pointing out that formal logic rationality does not seem to be the case by default, but also arguments leading to the conclusion that it is not by describing in complex ways the articulation of arguments in context that we necessarily end up with a faithful picture of how our minds deal with arguments.[148]

I will argue in chapter III that there are significantly powerful alternatives to start dealing with this question, notably stemming from the cognitive psychology and the cognitive anthropology domain. These two paradigms of research highlight the crucial role of cognition in argumentative practices, and call for further developments. The first is cognitive psychology. I have already mentioned past research on reasoning (cf. Wason 1966, Pohl 2004) and highlighted that the findings of such studies teach us that rationality and reasoning are not to be taken for granted. The second is evolutionary psychology and anthropology, and points to the importance of the cognitive underpinnings of argumentation in the evolution of the species (cf. Mercier & Sperber 2009, Sperber *et al*. 2010, Mercier 2009). Notably, the notion of epistemic vigilance, conceived as a set of screening devices that evolved to make sure that the information that is communicated to us is filtered so as to ensure its reliability constitutes an interesting focal point for the study of argumentation. It is with that paradigm in mind that the first lines of a cognitive model of manipulative and deceptive communication will be sketched out further along.[149]

The preceding sections were meant to assess some of the insights approaches in discourse analysis and argumentation theory have to offer in an account of manipulative and non-cooperative communication. Focusing first on the notion of discourse and on its structure and components in ordinary conversation, be it argumentative or otherwise, I have tried to show how these have attempted to account for the way influence could be exerted on language users,

[148] Yet, it should be noted that such is not the goal set by informal logic.

[149] See Chapter III.

notably through an analysis of the characteristics of the message, which calls for careful examination of underlying ideologies, the way these are brought about by discursive strategies and also argumentative moves, thus contributing to a study of influence centred on discourse, taken somehow as an object of its own, independently of the cognitive implications of discourse processing. Quite dominantly, hence, these approaches are message-centred, insofar as the message itself represents the starting point of an analysis of how words may impact beliefs, and thus influence audiences.

In the following section I bring the discussion to an extra-linguistic level by looking in more detail at the type of constraints external parameters may exert on behaviour.

II.2. CONSTRAINTS ON BEHAVIOUR

II.2.0. INTRODUCTION: INFLUENCE IN SOCIAL PSYCHOLOGY – ITS LINK WITH THE STUDY OF MANIPULATION

The present sub-chapter will review some major works in social psychology that expose how behavioural influence can be achieved 'from the outside', that is, through mechanisms which hinge upon social constraints. I will thus list some parameters which may also constrain what language users make of communication, but this time from an external perspective. In doing so, I will not spend excessive time on a critical evaluation of the experimental and theoretical aspects of this research, as my main goal in this section is to point to specific evidence for particular aspects of human interaction which I will draw on later on as I reinterpret some findings of argumentation theory and discourse analysis within a cognitively grounded model of manipulative and non-cooperative communication.

Many trends in social psychology aim to uncover how particular behavioural constraints which are imposed onto the individual from her/his social environment may lead her/him to either engage in specific courses of action or adopt specific beliefs; these are very valuable in understanding why it is that influence can work. Nonetheless, little has been done in the past to systematically link this body of research to cognitive aspects of communication conceived more broadly in order to see how these constraints may be realised in discourse, and whether there are specific discursive strategies allowing to gain insights into the workings of influence through communication.

The study of influence has gradually emerged as a paradigm of study in social psychology in the course of the 20th Century, with researchers' increasing interest in inquiring about the conditions under which social relations are decisive to explain individual behaviour. Many studies (the most famous of which are probably Milgram's (1974) and Asch's (1955), see §II.2.1.1. and §II.2.1.2. below) were explicitly designed to test some of the social variables that were assumed to play an important part in determining behaviour, and in particular in assessing how the influence of a given group on its members is regulated. While the amount of variables one has to take into account when trying to study human behaviour is extremely high and tentatively difficult to pin down, the experimental results reached within this paradigm are characteristically confirmed through abundant replication, and thus the phenomena they bring to the fore can reasonably be taken to reflect actual behavioural tendencies.

What is interesting, as far as these experiments are concerned, is that their results can also be interpreted as having implications for the description and explanation of specific communicative phenomena such as deceptive argumentative strategies, insofar as there are solid grounds to

consider that the efficiency of these is underlain by the very mechanisms uncovered by social psychological research. In particular, I will contend that many – if not most – argumentative fallacies have such a psychological counterpart (for illustrative purposes I will evoke later on the *ad verecundiam*, the *ad populum*, and the *ad hominem*), meaning by that that fallacies work because they take advantage of the cognitive features of our information-processing mechanisms.[150]

The present section will thus review the type of research that is relevant for such an account. I will first pay attention to the social strand of research in social psychology, before moving on to the paradigm that devotes more attention to the psychologically-oriented study of the cognitive underpinnings of social influence.

II.2.1. SOCIOLOGICALLY-ORIENTED SOCIAL PSYCHOLOGY

II.2.1.1. Milgram's experiment

Social psychologist Stanley Milgram has gone in history as the designer of perhaps the most troublesome experiment of all time. Shocked by the unquestioning way in which soldiers executed orders in WWII, leading to the murder of millions of people, he decided to undertake research into the mechanisms of obedience in an attempt to explain how people may end up perpetrating actions contravening their moral standards under specific social conditions of irresponsibility and through the intervention of authority-exerting figures.

Here is how his experiment would go. Milgram would first recruit subjects who were told they were needed for an experiment of memory (that is what the announcement said), and they were told upon their arrival that the goal of the experiment was to study how memory and learning could be influenced by punishment, the punishment consisting in the administration of gradually increasing electric shocks. Two subjects and one experimenter would participate in the experiment. After agreeing on the terms of the experiment, the subjects would draw lots in order to determine their role; what the naïve subject – the one actually being tested – did not know is that both papers read 'teacher', that the other participant was a confederate of the experimenter's who was instructed to say that he picked the paper with 'learner' written on it, and that this was a ruse meant to make sure that the naïve subject would always end up being the teacher. Both participants were then taken to the room where the learner would sit on a chair, the armrest of which was connected to an electric-shock-dispensing device. The experiment would take the teacher's arm and apply the device onto it in order to show him the

[150] See Oswald (2010).

kind of shock that the learner would receive, of course, with a very light voltage. Having attached the device to the learner, the experiment and the teacher would then leave the room to go to another room with the electric-shock-generating machine comprising several switches corresponding to increasing voltages ranging from 15 to 450 volts, this last level being labelled "XXX" on the machine. The teacher was then instructed to read a list of word pairs to the learner, who had to remember them. After that, the teacher read the first word of each pair and then offered four possible solutions from which the learner had to pick the correct item in order to match the. Any wrong answer had to be sanctioned by an electric shock administrated by the teacher, who would also announce the voltage each time. Every new mistake meant an increase in the voltage (15 volts at a time). Pre-recorded reactions to electric shocks ranging from light to heavy and painful screams were played in the other room, so the teacher could hear what he thought the learner's reactions were. As the voltage increased, reactions were becoming more and more disturbing, leading the teacher to believe that the learner was experiencing too much pain. Past a certain threshold, no reaction at all would emanate from the room where the learner was sitting. None of the electric shocks were obviously real, but the teacher was led to believe that such was the case, and this was crucial for the real object of study, i.e. a subject's obedience to authority.

It was expected that many tested subjects would want to stop the experiment as the screaming coming from the other room became more and more unbearable. Thus, the experimenter was instructed to try to get the teacher to proceed. A routine of four invariable answers ("Please continue"; The experiment requires that you continue"; "It is absolutely essential that you continue"; "You have no other choice, you must go on", cf. Milgram 2004[1974]: 23) was systematically applied, and the experiment would stop only if after these four reactions on the experimenter's part (who was wearing the characteristic grey blouse of the scientific researcher) the teacher would still want to stop. The experiment was also brought to an end after the teacher delivered the highest voltage (450 V) three times in a row.

An astounding 65% of subjects pressed all the switches, from a mild shock to a potentially lethal one. This dramatically contrasted with the results of the polls Milgram had obtained from previous interrogations of fellow psychologists and colleagues, who predicted that less than 2% of the tested subjects would go all the way. Replications of the experiment, varying physical and geographical location, gender and other parameters (see Milgram 2004[1974]) confirmed the results obtained in the original trials. However, some parameters were shown to decrease obedience, such as victim proximity (in one condition the teacher had to physically hold the hand of the learner and press it against an electrified plate), experimenter's contradiction by another experimenter, or replacement of the experimenter by an ordinary man.

In order to account for these results, Milgram postulated the existence of an *agentic state* which is requires an individual to be in a hierarchical relationship, to turn to someone hierarchically above her/him for guidance and to usually seek to transfer the responsibility of her/his actions to someone else's:

> "The essence of obedience consists in the fact that a person comes to view himself as the instrument for carrying out another person's wishes, and he therefore no longer sees himself as responsible for his actions. Once this critical shift of viewpoint has occurred in the person, all of the essential features of obedience follow." (Milgram 2004[1974]: xviii)

The idea is that the subject experimenting a tension between two constraints would delegate his responsibility to that of the experimenter's, thereby reducing the strong discomfort that would be caused should s/he endorse her/himself the whole responsibility of inflicting pain onto someone else.[151]

It is not my point here to delve into the discussion about the proper explanation of obedience to authority.[152] However, the fact that the experiment has been replicated numerous times[153] should at least tell us that there is such a thing as reliance on (and obedience to) authority in human behaviour. In other words, it could very well be the case that trusting an authority appears to be a natural thing to do when one is unable (be it by virtue of ignorance, incompetence or moral discomfort) to take a decision in a given context. The crucial aspect that I would retain from Milgram's research lies in the idea that obedience to authority is precisely what some argumentative strategies are designed to achieve. The argument from authority (see van Eemeren & Houtlosser (2003), for instance), or *argumentum ad verecundiam* in its fallacious form, precisely tries to make a standpoint acceptable by positing that a given authority (illegitimate in the case of the fallacious move) holds that standpoint as acceptable. In other

[151] Note here the similarity with the type of double-bind Festinger (1957) has in mind with respect to his notion of cognitive dissonance: either solution is disadvantageous/uncomfortable in at least one respect, which is why a state of cognitive trouble is induced.

[152] For an interesting alternative explanation taking into account the cognitive implications that the belief in the reasonableness of science may have in this experiment, see the brief note by Nissani (1990), who opens up the debate to cognitive considerations; he points out that in order to disobey in the Milgram experiment, subjects needed to give up their rational assumptions about the responsibility and morality of the scientific design of the experiments they were taking part in. This in turn suggests that additional cognitive parameters having to do with previous beliefs, and not solely obedience to authority, have to be taken into account in the explanation of the phenomenon.

[153] Milgram's book collects 18 replications tweaking several parameters. In other fields, such as television, the experiment has also been replicated, even very recently on French television ("Le jeu de la mort", 2009, see also footnote 159). All these variations nearly systematically yielded the same results. People are prone to obeying authority, even when this means inflicting pain onto another human being.

words, the reason why this type of argument might very well be effective could be explained by our natural favourable dispositions towards authority – or at least expertise. It seems we trust knowledgeable people and it turns out to be perfectly rational: when we do not know something, a safe bet is to go with whatever those who know better (or at least those we think know better) think and say.

Also elaborating from social psychological research, Cialdini (2009) points out that the substantial (i.e. actually verified) authoritative status of someone may not be the only cue underlying our disposition to obey or conform to what those we believe to be experts tell us. Various symbols can make up for an *appearance* of authority, including clothing, titles and cars (see Cialdini 2009: 184-197), and this may be enough to trigger the expected behaviour. This in principle opens up the possibilities for people attempting to play the authority card in order to influence an audience; typically, the mention of any element purported to contribute to the identification of an individual as knowledgeable, wise, expert, and so on, is potentially an effective means of eliciting compliance. These are the main reasons why Authority is one of Cialdini's (2009) six principles of influence.

Evidence stemming from a first-hand source as far as deception is concerned, Kevin Mitnick (a former social engineer who has hacked computers and penetrated security systems for years), comes to illustrate this principle: sometimes claiming to be an authority or appearing to be a reliable and knowledgeable person is sufficient to gain information and more generally benefits – benefits that could be highly detrimental to someone else or to a corporation. Anyone calling an employee, claiming to be a telecom technician that needs the employee to perform certain 'maintenance' operations on one of the computers of the company, and who furthermore manages to appear as a legitimate expert (by knowing the lingo or by resorting to accurate name-dropping about company employees) might have a chance to influence the person on the other side of the phone line so as to gain access to classified information. This is one of the basic strategies social engineers use to get things their way (see Mitnick 2003).

A social constraint such as obedience to authority provides an empirically grounded explanation of why certain mechanisms of influence might work, and as such they are legitimate and potentially enlightening for the account of manipulation and uncooperative communication I am developing in the present thesis. This constitutes a first inquiry into the underpinnings of the *ad verecundiam* fallacy; what this particular fallacious move does in an argumentation is precisely to point out that reliance on authority has a social reality because people have experienced it

previously in their lives.[154] I will return to this particular case in Chapter III when I reinterpret these findings in terms of information-processing, as I try to show from a cognitive viewpoint how exactly reliance on authority can sometimes mislead people. But for the time being, let me proceed with additional experimental evidence showing that some specific arguments actually map in discourse what social psychologists have found on the level of behavioural influence.

II.2.1.2. Asch's experiment on conformity

Solomon Asch, who was actually Milgram's supervisor, worked on another aspect of social interaction and influence that eventually came to be known as the underlying mechanism of peer pressure. He conducted an experiment (reported in Asch (1955)) where groups of students were asked to perform what was presented as a vision test: participants had to judge which of three uneven lines was equal in length to a line which they were presented first as the standard measure. These lines were presented on cards (see figure 2).

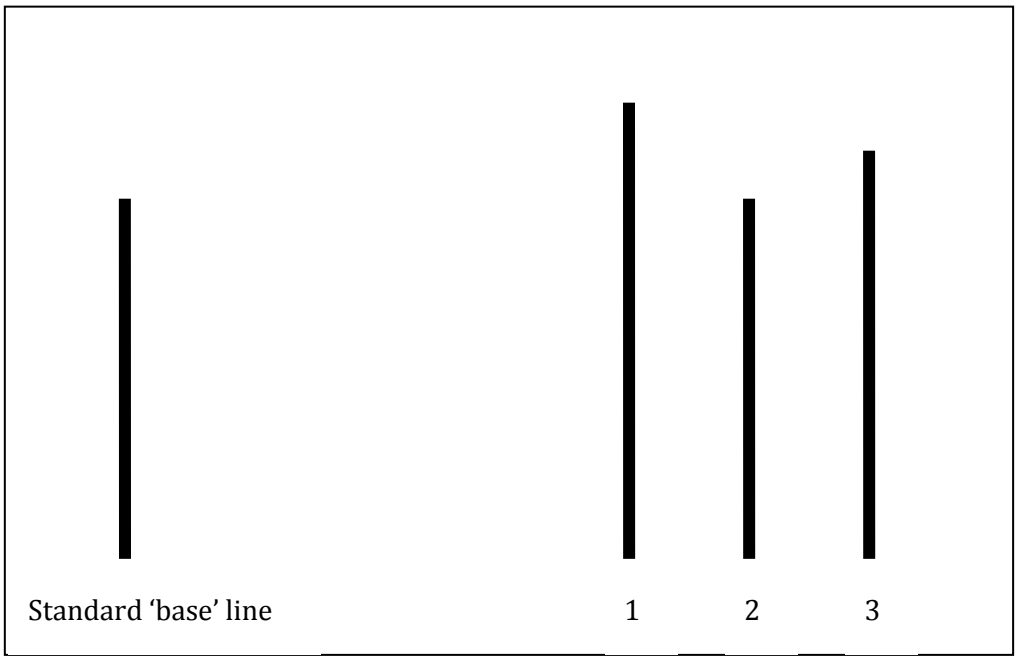


Figure 2. Material of Asch's experiment

[154] Maillat (2010) explains that although fallacies do not actually perform some of the effects we think they hinge upon, they instantiate somewhat 'discursively' something that we have had experience of in the past, and they act as triggers of the intended effect. With respect to the case in point, the *argumentum ad verecundiam*, the fallacy may not effectively be about obeying a present and real authority, but makes salient that there is an authority in the context, whose opinion is decisive in what regards the standpoint or the task at hand.

In reality, Asch was testing only one of the participants' disposition to produce judgements that conformed to what other participants said despite perceptive evidence; unbeknownst to the naïve subject, all other participants were confederates who had been instructed beforehand to unanimously provide wrong answers at specific times. So, while the first rounds would usually pose no problem at all and yield unanimously convergent responses, later rounds generated situations where only the naïve subject would be inclined to pick, say, line 2, according to what he saw, while all the others picked line 3 (see figure 2). What is more, the naïve subject would answer after the others, so he was in a position where he could compare other participants' answers to his own perceptual impressions. What Asch observed is that 37% of the naïve subjects (out of 123) went for the (erroneous) answer given by the rest of participants, thereby contradicting their own perceptual evidence.

Although the effect is not as robust as Milgram's in terms of resulting percentages, it does tell us something about the role knowing what others do and think can play in shaping our behaviour. A critical point should nevertheless be raised here and has to do with the nature of that which the experiment tells us social pressure can influence. The results show that 37% of people produce an answer which does not conform to their own perception but to what others are saying. But this is quite different from saying that 37% of people end up *actually believing* that the line they initially picked is in fact different from the base line. In other words, just because 37% of subjects pick a line out of conformity in no guarantee that they actually believe that they were themselves mistaken in the first place and that they believe that whatever their peers say is correct. Hopping on the bandwagon does not necessarily imply honestly expressing true enjoyment for the music played. Social and face-related motivations, as far as consequences of disagreement are concerned, are distinct from epistemic matters in the sense that one can engage in a given course of action without having to intellectually endorse the epistemic consequences of it, but this is not something that the data of Asch's experiment can explain straightforwardly. Put differently, the effect of conformity may explain behaviour but not necessarily belief.

Ciladini's theory of influence identifies 'social proof' as one of the six principles of influence. When we do not know what to do, or when we find ourselves in situations which we do not know how to deal with, it seems that we have a natural tendency to observe what others do and take that as a reliable measure for appropriate behaviour. This is how Cialdini explains in turn many phenomena. For instance, this is according to him the reason why bartenders leave a few coins or one-dollar bills in the tip jar *before* they start their shifts. The idea is that customers will be influenced into leaving coins because there is evidence that other people did it too before them, and thus also because this is taken as proper behaviour under those circumstances. This

mechanism is also behind a particular scam taking place nowadays in the streets of some cities, known as the 'Three Card Monte'. In this scam, a man shuffles three cards face down, two of them being spades and one hearts. The passer-by who gets caught in the game must then pay a (usually large) sum of money and tell the shuffler where he thinks the card of hearts is.[155] Usually there is more than one man running the scam: there is the shuffler, watchmen looking out for police officers, but more importantly an accomplice pretending to be a random passer-by just like any other; this individual plays, and wins at least twice so that people around are encouraged to believe that winning money in this game is actually easy (since they would have gone for the same card as the apparently random passer-by). Of course, when the innocent target decides to play, the shuffler puts in practice his real skills and manages to fool his victim. It could be argued that what prompts the victims of this scam to play is the fact that they have seen other people easily win. In other terms, they could be assumed to follow the 'if they won easily, then why wouldn't I as well?' reasoning pattern, which is in turn vivid illustration of the mechanism Cialdini refers to as the *social proof*.

Experiments by Bandura (cf. Bandura *et al*. 1967) also showed that the social proof principle can motivate radical changes in behaviour, including desirable ones. He made nursery-school age children who were particularly afraid of dogs and who would not come near one repeatedly watch other children play with a dog for 20 minutes at a time, and enjoying it. He then observed a change of behaviour in 67% of cases, what is more a change that would be preserved in time, the children showing more enthusiasm for playing with dogs later on. The behavioural pattern behind these experiments could very well be translated into some kind of inductive reasoning pattern: if many people do X, then X is acceptable/advantageous/moral, etc., and this constitutes sufficient grounds for me to do it too. Reviewing literature on the subject leads to the conclusion that conformity can be a source of influence, and sometimes a radical one.

Looking at the communicative side of the story, we cannot fail to notice that the similarities between what these experiments tell us on social pressure and the mechanisms underlying the success of the *ad populum* fallacy (also known as the appeal to consensus or the bandwagon fallacy) are striking. The appeal to consensus, whereby you may support a conclusion by resorting to the premise that many people endorse that conclusion, resonates with research on conformity and can be interpreted as a verbal instantiation of a mechanism which we have all experienced. Similarly to the *ad verecundiam*, the *ad populum* does not enforce its underlying mechanism, to the extent that one singularly occurring argument cannot be a proper repetition

[155] There is also a variant with upside down cups and a little ball which is passed on from one cup to another very quickly.

strategy; yet what the argument does is make salient a mechanism which humans have repeatedly experienced in their lifetime.[156]

II.2.1.3. Sociologically-oriented social psychology and manipulative communication

In what precedes I have described some rather famous pieces of research in mainstream social psychology which focus on the consequences external constraints might have on behaviour. As a field devoting much of its attention to the study of influence, social psychology has significant insights to offer for the description and explanation of behaviour, and this could be used to shed light on the issue of manipulative and uncooperative communication, to the extent that these may be also taken to act as constraints on behaviour. Yet, social psychology is far from being a homogeneous field of research, insofar as some of its proponents advocate a strong link with psychology, whereas others prefer to give precedence to sociological concerns.[157] In the latter view, behaviour is explained by considering the social conditions in which it is observed, therefore postulating the causal determinacy of social parameters. Within this perspective, behaviour is conceived as the fruit of social interaction and not of biological or cognitive processes. Psychological trends in social psychology, on the other hand, try to explore the relationship between both social and cognitive constraints in order to explain human behaviour with strong emphasis on the cognitive and psychological conditions without which behaviour would not be possible. Without getting into the nature *vs*. nurture debate, I will limit myself to highlighting that an intermediate position seems to be the safest bet (this is a position adopted by Clément 2006 for instance), since extreme positions on the issue are easily criticisable: *tabula rasa* theories just cannot explain cross-cultural behavioural regularities and exclusively psychological accounts just cannot ignore environmental influences on behaviour altogether.

Now, as far as communication is concerned, I have already hinted at the idea that fallacies can straightforwardly be exploited to fulfil manipulative purposes: a speaker who knows that the argument she is deliberately using is faulty and yet still tries to use it can be accountable for intending to manipulate her audience (which is not the case when the fallacy is produced unintentionally, and thus in good faith). Social-psychological accounts of the type of those I have evoked above are obviously the right theories to turn to if we want to assess psychological and

[156] See Maillat (2010) for the original formulation of the argument.

[157] This distinction is also found in adjacent disciplines: for instance, the distinction between sociolinguistics and psycholinguistics is reminiscent of the tension between social and psychological approaches to human behaviour.

social aspects of fallacious argumentation, in particular in what regards explanatory models meant to answer the question of knowing why fallacies may be effective (see Oswald 2010, Maillat 2010). This is something that will perhaps not inform us much about the structure of arguments *per se*, but it will surely be of great help when comes the need to explain why these arguments are convincing, thereby consolidating the rhetorical component of the account. What is more, they provide empirical grounds to backup more general assumptions about the way cognition functions, even more so should we want to assess how cognition evolved so as to be tailored to successfully manage standard communicative situations (see for instance Sperber *et al*. 2010, Origgi & Sperber 2000), which can in turn provide us with relevant insights to assess non-standard communicative situations such as manipulative and uncooperative scenarios.

Social psychology identifies conditions of behavioural influence, that is, it aims at explaining changes that can be witnessed in people's actions. Nonetheless, it may also be expected to explain changes in people's mental states, should we accept that changes in behaviour are motivated by changes in mental states, as advocated for instance by Michael Bratman in his 'Belief-desire-intention' model of practical reasoning which takes action to be a consequence of mental states such as beliefs, desires and intentions (cf. Bratman 1987 [1999])[158]. Moreover, it should be noted that social psychology has a great deal to say about cases of manipulation, since manipulation may typically involve strategies of influence, albeit of the 'illegitimate kind'.

Such a direction of research is nowadays very much alive in French speaking circles of social psychology research. The 1990's in particular have seen in that respect an increase in the rate of publication of books on influence in France, authored by researchers in the social sciences who took on Anglo-Saxon research of the XXth century in order to develop models of social change through influence (e.g. Paicheler 1985, Guéguen 2002, Chalvin 2003, Joule & Beauvois 1998, Bellenger 1997, Yzerbyt & Corneille 1994, Breton 2000). This pool of research usually chooses to tackle the issue of influence either in a functional way (by focusing on the uses of influence that are instrumental to fulfil 'good' and 'legitimate' social goals) or through the lens of its political and social significance, and thus it is usually driven by a pedagogical concern meant to enlighten readers so that they are better protected against social influence of the illegitimate

[158] See also Clément (2006) for an account of credulity which explains how beliefs, and thus behaviour, may be influenced. This account lies at the intersection of sociology, (cognitive) psychology and philosophy, which integrates both social and psychological influences on mental states. See also section §II.3.2.2.

kind.[159] The output of these publications is accordingly often intended to reach the general public and not exclusively the community of researchers.

Although they usually build on a varied range of works both in social and psychological social psychology, these approaches are usually driven by a normative political concern according to which influence of the 'bad' kind is ethically unacceptable; Breton qualifies manipulation in this respect as "a violent and constraining action that deprives from liberty those who are submitted to it" (Breton 2000: 23, my translation) and thus positions himself in the ethical debate as a proponent of an emancipatory practice of scientific research.[160] His idea – close to Chomsky's, who claimed that propaganda is to democracy what violence is to the totalitarian state – is that manipulation is detrimental to the practice of democracy; consequently, fighting manipulation should be a driving force of research in social psychology, since the latter is characteristically the field of research devoted to the study of behavioural influence. Similarly to CDA's (see §II.1.1.), the political agenda of the type of approach defended by Breton runs the risk of leading to some carelessness in the analysis. For instance, he claims that "there is manipulation because the reason that is given for adhering to the message has nothing to do with the content of the message; it is precisely through this property that we are technically able to identify it" (Breton 2000: 80, my translation).

This assertion is probably too radical, for two main reasons: first, sometimes people provide reasons which are not relevant to the content of the message, yet without intending to manipulate the addressee; after all, people do make mistakes and errors in reasoning, and just because this is reflected in the way they phrase their thought is no guarantee of there actually being manipulation at all. Second, there are also cases of manipulation where the speaker issues a statement which is fully cooperative in terms of truthfulness and sincerity (see example given

---

[159] We can quote for illustrative purposes a very recent replication (2010) of the Milgram experiment which was conducted in a TV-game setting ("The game of death", documentary directed by Christophe Nick), with the participation of Jean-Léon Beauvois' team of researchers in social psychology. One of the outcomes of the documentary was to reflect upon the power of television (which in my opinion is a false problem since there is no way in current political and legal system that someone might be knowingly injured or killed on TV; the documentary by no means demonstrated that television is powerful enough to lead to voluntary torture or murder), but independently of that, the experiment continued to yield the same consistent effects with respect to the role of authority in mechanisms of influence. Ultimately, it seems that the whole show was conceived as a pedagogical sensationalist method of raising people's awareness against the dangers of blind obedience to authority (in this case a TV host and the audience, who would play the part played by the experimenter in Milgram's original setting).

[160] In this respect his approach converges with CDA's agenda (see §II.1.1.), since research under this perspective is instrumental to the goal of protecting people of something by letting them know how that something operates.

in §I.2.3.2.1. above and also in Maillat & Oswald forth.). The idea of linking the account of manipulation to ethics and politics is not to be abandoned altogether, but it seems to me that if we are to adopt a naturalistic take on these phenomena, aiming to explain how these work from the perspective of information-processing, we need to exclude from the scope of research the moral and ethical aspects of manipulation in order to gain knowledge on the reasons that make for its success.

All and all, the kind of research developed by socially-oriented social psychology is of great value to an account of influence. However, as it focuses on the external constraints of influence, it usually neither focuses on communication proper, nor on its cognitive underpinnings. I have already mentioned earlier that there are linguistic/discursive counterparts to the sort of influence labelled as 'obedience to authority' or 'social proof', which may take the form of *ad verecundiam* and *ad populum* fallacies respectively. From this kind of analogy it should be apparent by now that the type of research carried out in social psychology is of great interest for the study of manipulative communication; nevertheless, as social psychology of the sociological kind has not specifically addressed the communicative specificities of such mechanisms, it makes sense to try to integrate its results into a communication-oriented account of manipulation such as the one I am developing in the present thesis. Crucially, this supposes that an explanation of social facts should be grounded on cognitive considerations; as noted by Sperber, "the thread that links all social events is a cognitive thread that goes through minds, through the environment, through minds, through the environment, and so on, securing content relationships along the way" (Sperber, forth.). The underlying idea of this methodological stance is the central assumption that social facts would not be possible in the first place if our cognitive machinery was not equipped to deal with them.

II.2.2. PSYCHOLOGICALLY-ORIENTED SOCIAL PSYCHOLOGY AND MANIPULATION

I have mentioned earlier that some strands of research labelled as social psychological have extensively dealt with cognitive mechanisms of persuasion. Although these are concerned with social behaviour, they indeed pursue research directions of great interest for scholars interested in the cognitive underpinnings of influence, be it legitimate or illegitimate.

As pointed out by Clément, this strand of research promotes the study of *attitude*, which rests on the idea that we can introduce "between the stimulus and the response a representational mediation whose function is to provide an evaluation of the stimulus" (Clément 2006: 111, my translation). Attitudes are therefore taken to refer to the dispositions an individual might have about objects, people, events and institutions, in sum, basically about anything that can be represented and evaluated. Two main aspects of attitudes are relevant to social psychology:

their evaluative character (positive and negative) and, as noted by O'Keefe (2002), the fact that attitudes and behaviour are usually consistent.[161] That is, reaction to a stimulus will vary according to whether the stimulus is favourably or unfavourably welcomed by the individual; favourable attitudes will most probably yield favourable reactions, while unfavourable attitudes will be expected to yield unfavourable reactions. This of course bears consequences on the individual's behaviour, but additionally, attitudes will also affect the individual's beliefs (Clément 2006). And it has a more interesting consequence as well: if someone wants to have things their way, then they should try to trigger favourable attitudes towards the object of their strategy of influence. These elements converge towards one conclusion: gaining knowledge about how attitudes work is crucial to gaining knowledge about the underpinnings of both legitimate (avowable) and illegitimate (covert) influence.

II.2.2.1. Persuasion research: the Elaboration Likelihood Model

II.2.2.1.1. General overview

Scholars within the field of persuasion research have devoted much of their attention to studying attitudes and the conditions under which attitudes are likely to change.[162] Perhaps the most famous are Richard Petty & John Cacioppo, who designed a theory meant to account for the way persuasion works from a cognitive point of view, namely the Elaboration Likelihood Model (henceforth ELM).[163] ELM posits that individuals presented with a persuasive message may process it in different ways, ranging from what they call high to low elaboration. High elaboration refers to the amount and intensity of processing they devote to thinking about the issue at stake (otherwise named amount of 'issue-relevant thinking'). This is typically what happens when one engages in reflexive and argumentative reasoning where careful attention is paid to the message, its consequences, implications and relation to other previously held information, and characterises a type of processing that one could call 'committed'. This type of processing is referred to as 'central processing'. At the other end of the spectrum, the nature of processing of information that is meant to persuade can also follow a path of least effort, usually deemed intuitive, non-reflexive, and heuristic-based. This is called the 'peripheral' route to persuasion, and under these conditions, the mind is taken to rely on simple rules of thumb or on

[161] Even if this does not mean that attitudes and behaviour are always consistent; O'Keefe highlights for instance that we may have a positive attitude towards good health, and yet still fail to adopt a behaviour consistent with that attitude, such as eating less fast-food, doing more exercise, quit smoking, etc. (cf. O'Keefe 2002: 20)

[162] See for instance Audi (1972), Breckler & Wiggins (1989), Eagly & Chaiken (1993), McGuire (1985).

[163] The following discussion draws on O'Keefe's informed exposition of ELM (O'Keefe 2002, chapter 6).

specific heuristics that make processing less effortful than required by the central route to persuasion. Here typically, the message will not be processed with careful attention, and only some of its superficial properties will be accessed (such as trustworthiness of the source, resemblance of the form of message with other forms, which can in turn very quickly lead to acceptance of the message's content).

The presence of two apparently radically opposed modes of information-processing is neither evidence that they are mutually exclusive nor evidence that they cannot intervene both in the processing of the same persuasive message. ELM posits that there is actually a continuum between both extremes and that it is conceivable to postulate processing of persuasive messages that taps into both routes at the same time, albeit in variable proportions (see O'Keefe 2002: 140). What this theory shows is that the intended outcome of a persuasive message, namely the adoption of a certain belief or behaviour by the addressee, can result from cognitive processes of different natures. Additionally, this supposes that the factors relevant to persuasive success depend on the route processing takes.

Central processing is believed in ELM to be affected by the addressee's motivation and ability to process the message. Motivation is determined by personal relevance of the message (i.e., whether the message carries strong implications for the addressee's general well-being, or whether the message is particularly important with respect to the addressee's goals, desires and intentions), but also by the addressee's need for cognition, defined as "the tendency for an individual to engage in and enjoy thinking" (Cacioppo & Petty 1982: 116), which varies from one individual to another. Thus, the more the issue is personally relevant to its target and the more eager to engage in thinking in general the individual is, presumably the more involved the latter is going to feel in processing the message.

Ability to process the message is more or less defined as the material possibility to do so; factors such as distraction and previous knowledge about the issues relevant to the message have accordingly been studied in order to determine their impact on one's ability to elaborate the message. The main hypotheses that have been tested in relation to this last point establish that addressees holding unfavourable attitudes towards the message end up more easily persuaded when distracted by a parasitic stimulus, but also that addressees who are favourably disposed towards the message do not end up persuaded when their ability to engage in issue-relevant thinking is decreased by means of distraction. As far as prior knowledge is concerned, it is believed that informed addressees are likely to engage in thorough issue-relevant thinking and consequently that persuasive success increases as the target is able to elaborate on the topic.[164]

[164] See O'Keefe (2002: 137-167).

Studies in ELM tend to show that as elaboration increases, the importance of peripheral cues diminishes, and thus that individuals are likely to engage in central processing rather than peripheral. In this case, the central route is the path that will lead to persuasive success. Conversely, as peripheral cues become more important (that is, as the addressee is prevented from elaborating the message at length and in great detail), central processing tends to be secondary, and in those cases persuasion will result from heuristic-based processing of the message. The degree of elaboration is thus subject to different factors, but studying these, as we have briefly seen, is only one of the aspects of persuasion research. The other, crucial, aspect of this strand of research is to assess which factors determine persuasive effectiveness, both in central and peripheral processing of persuasive messages. This is something referred to as 'elaboration valence'. The idea is that individuals are persuaded by a message when the type of processing they engaged into lead them to entertain favourable attitudes towards the message; favourable attitudes will ensure persuasion, unfavourable ones will hinder it. The question that ELM addresses from that point on, then, is that of uncovering the parameters orienting attitude (for both routes).

Central processing, which encourages intensive issue-relevant thinking, is believed to be affected first by the addressee's initial attitude towards the content of the message. Arguably, people who already entertain favourable attitudes towards a message which advocates information which is consistent with these attitudes will be persuasive; in parallel, people who already entertain unfavourable attitudes towards the content of the message will not be persuaded by a message which is inconsistent with these previously held attitudes (such as counterexamples and counterarguments). But this is not sufficient to explain persuasion, since, as O'Keefe observes, this would mean that "nobody would ever be persuaded by a counterattitudinal message" (O'Keefe 2002: 146). An additional parameter, dubbed 'argument strength' or 'argument quality' has thus been studied, with the underlying assumption that strong arguments contribute to persuasion in high elaboration scenarios. Yet, as noted by O'Keefe, the way argument strength is defined is problematic, since ELM does not provide criteria of argumentative quality, which leads to an inoperative circular definition whereby strong arguments are the ones that are persuasive. From these elements, it seems that what ELM has to teach us about the parameters prompting for high elaboration of the persuasive message appears to be quite limited; indeed, it what is missing from the account is more research about the actual parameters playing a role in leading the addressee to engage in such effortful processing.

II.2.2.1.2. Discussion

The peripheral route of persuasion uses heuristics and more general rules whose application does not require a great deal of elaboration; as such they are mostly unconscious and automatic, which makes them harder to study, to the extent that the experimenter cannot question subjects about them and expect reliable answers.[165] Persuasion research has studied some of these heuristics, like the credibility heuristic (according to which you are more likely to favourably welcome a message whose source you deem reliable), the liking heuristics (according to which individuals are more likely to be persuaded by speakers they like)[166] and the consensus heuristics.[167] I will not get into these experimental designs here, as I will deal with related aspects further along (see §II.3.1.2.). However, there is a first hypothesis to be drawn from the existence of these heuristics: as peripheral information-processing is an alternative to high elaboration of the message, speakers who are driven by manipulative intentions are probably likely to try to get their audiences to rely on these heuristics, since high elaboration would potentially allow them to infer information which is detrimental to the manipulative intent (such as inconsistencies that are internal to the argument, inconsistencies between the content of the persuasive message and the addressee's own values and beliefs, etc.). This is nevertheless to be taken with great caution, for sometimes subjects who are explicitly instructed to look for anomalies in a message still fail to spot them, that is, even when subjects are encouraged to engage in issue-relevant thinking, they still might be fooled.[168] I will come back to these issues shortly in the section devoted to cognitive psychological research (see §II.3.).

An interesting finding of ELM is the idea that people's processing of information can be influenced by different parameters, and notably the idea that there are actually different ways involving more or less cognitive effort to reach the same result (i.e. being persuaded). As such, ELM and similar theories (such as the Heuristic Systematic Model, cf. Chaiken *et al*. 1989) represent an appealing field of study to start assessing the conditions that make for persuasive effectiveness. By extension, we could even consider, following for instance Clément (2006), that

---

[165] This however is not the case with central processing testing: one of the experimental methods to investigate high elaboration is to ask subjects to self-report what thoughts occur to them as they process the stimulus, which allows representing the reasoning pattern.

[166] Notice the evident parallel with Cialdini's' 'liking' principle (2009).

[167] Notice as well that here we find again a striking similarity with what *ad verecundiam* and *ad populum* arguments exploit.

[168] See for instance Barton & Sanford (1993) who show in the so-called 'Moses illusion' paradigm that despite being warned about potential anomalies in the narrative presented in the experiments, that is, despite implicitly being motivated to process the information systematically, subjects still fall victim to the lexical anomaly by failing to spot it. See also section §II.3.1.3.1. for an elaboration on the Moses illusion.

such research has a great deal to offer to the study of manipulation, should we consider that one of the necessary conditions of manipulation is the covertness of the manipulative intention (see §I.3.1.3.): in this respect, trying to prompt the addressee to rely on heuristics, or so-called cognitive shortcuts, more than elaborate thinking when processing the message is actually a way of leading him to miss certain information that would be detrimental to the manipulative attempt. As noted by Clément,

> "peripheral processing of information […] is particularly likely to sustain manipulation. […] Skilled manipulators can thus stage situations were misleading clues will be likely to trigger one of these shortcuts, which will in turn make their message more believable to the eyes of their 'targets'" (Clément 2006: 129-130, my translation).

This last point opens up promising directions of research; some have actually partly already started to be explored (cf. Maillat & Oswald 2009, forth.); the latter clearly resonate with a few findings typical of persuasion research. Notably, ELM posits that the degree of elaboration invested in processing the message depends on certain factors such as motivation and ability; yet, what could be said in addition to that is that the degree of elaboration might also depend on the (un)availability of cues contained in the message and which prompt for either central or peripheral processing of the stimulus. That means that some messages will make it relevant to engage in central processing, while others will privilege heuristic-based treatment (I will return more extensively to this idea in Chapter III).

II.2.2.2. Festinger's theory of cognitive dissonance

II.2.2.2.1. The theory

The theory of cognitive dissonance, elaborated in the 1950s, is one of the first attempts at systematically explaining the link between behaviour and so-called *cognitions*, which are defined by Festinger as knowledge, beliefs or opinions about the world or oneself – but which we could nowadays probably interpret more globally as mental states. Festinger's definition in fact leans towards this direction, only he restricts the notion to that of knowledge to define cognitions:

> "These elements [the elements between which a relation of consonance or dissonance may exist] refer to what has been called cognition, that is, the things a person knows about himself, about his behaviour, and about his surroundings. […] Some of these elements represent knowledge about oneself: what one does, what one feels, what one wants or desires, what one is, and the like." (Festinger 1957: 9)

Festinger expresses the thesis according to which action – or behaviour – is usually consistent with cognitions. That is, extending the scope of the term cognition so as to include all kinds of mental states such as intentions, but also desires and goals, we assume that the mind is such that

it strives to maintain some kind of inner equilibrium or harmony by trying to avoid inconsistencies. For example, it would be very costly and difficult for someone to hold two contradictory beliefs at the same time: I cannot believe that Switzerland is located in Europe, and at the same time entertain the belief that Switzerland is not located in Europe without experiencing some kind of epistemic discomfort. Furthermore, the tendency towards avoidance of inconsistencies is not solely directed toward beliefs; Festinger observed that most of the time people also try to maintain consistency between cognitions and actions.[169] The state of consistency among cognitions or between cognitions and actions was termed *consonance* by Festinger. The state of inconsistent cognitions or cognitions and actions is accordingly referred to as *dissonance*: "*x* and *y* are dissonant if not-*x* follows from *y*" (Festinger 1957: 13) The mind, as a cognitive system, is hence believed to strive for consonance and try to reduce or avoid dissonance whenever (potential) dissonance emerges, that is, there would be a natural pressure to reduce dissonance when encountered.

The canonical example of cognitive dissonance is that of the smoker, who, in spite of having learned that smoking is dangerous for one's health, continues smoking; in this case he has cognitions about smoking which are incompatible with his behaviour, but he may also be said to entertain dissonant cognitions, in so far as two cognitions about smoking he may have, such as "I like smoking because it makes me feel good" and "smoking might kill me" are themselves dissonant with respect to each other. The nature of dissonance need not be exclusively logical; two cognitions may be dissonant indeed because they are logically contradictory, but dissonance can also emerge as a consequence of cultural standards (this happens for instance when one knowingly does something which is forbidden by politeness standards in a given community), opinion embeddedness (Festinger (1957: 14) gives the example of a Democrat voting for a Republican candidate which he happens to prefer: this generates dissonance not by logical standards, but because "I should vote for a Democrat" should follow from "I am a Democrat"), or past experience (if I know from experience that long exposure to the sun will burn my skin and that I stay out for two hours one sunny summer afternoon without realising that two hours have gone by, the cognitions associated to those two beliefs can be said to be dissonant).

Dissonance can vary in magnitude, generally depending on the extent to which the relation between two given cognitions is made salient, but also, as highlighted by Festinger, depending on the importance of the cognitions involved to the individual; someone may experience moderate dissonance, for instance, if they know that the grocery store closes in 10 minutes but

[169] Even if it is true that such is not always the case, any cognitive system is in principle geared towards consistency, given that the requirements of maintaining a system filled with inconsistencies would presumable be too great to be handled by our mental abilities.

out of laziness lacks the motivation to go there to buy something to eat (and knowing that other options will still be available for supper). Greater dissonance may be experienced when the issues at stake are of great importance to the individual: for instance, a tenant with both money and dental problems might have to choose between paying the rent or getting their teeth treated, knowing that failure to pay the rent drastically increases their chances of losing their flat. What is interesting is that the magnitude of dissonance between all the elements relevant to a given question is defined by Festinger in terms of relevance: "the total amount of dissonance between this element [i.e., the given question] and the remainder of the person's cognition will depend on the proportion of relevant elements that are dissonant with the one in question" (Festinger 1957: 17). This means that the relevance of certain cognitions to an individual will determine the degree of dissonance; if many relevant cognitions are dissonant with a given behaviour, then dissonance will be high, this in turn putting the individual under pressure to try to reduce the experienced dissonance.

There are several ways in which dissonance can be reduced. First of all, when experiencing dissonance about a specific issue, the individual can try to bring in as many issue-consonant cognitions as s/he can, so that their amount outnumbers the amount of issue-dissonant cognitions. This is typically what we can achieve by *a posteriori* rationalisation; with respect to the example of the smoker, we could imagine that the smoker brings in information such as "my Grand-Dad smoked until he was 90, and he was fine", "smoking reduces my stress", "smoking is actually good to prevent a cold", etc., which are all consonant with the actual act of smoking. An additional strategy consists in reducing the amount of issue-dissonant cognitions, and this can be achieved simply by refusing to think about them and by trying to drive them away from the horizon of consciousness; focus of attention is therefore driven onto consonant cognitions. This is something that can be achieved for instance through emotional reactions; Clément's notion of *emotional filter* (Clément 2006: 223-228) is particularly insightful in this respect. He postulates that self-deception may be explained by the individual's unconscious and automatic propensity to keep cognitions which are detrimental to the organism's well-being away from the focus of attention, so that they are readily disposed of. In terms of cognitive dissonance theory, this could be one way of reducing dissonance. Another way of reducing dissonance would be to assign a higher degree of likeliness to issue-consonant cognitions, and a lower degree of likeliness to dissonant ones; this is typically something that could in principle be achieved by relying on expert opinion, consensus or comparison (which, as I have already noted, can be achieved in communication with informal fallacies such as the *ad verecundiam*, the *ad populum*, or more generally any fallacy aiming to ensure focus on one proposition over alternative others).

All in all, it seems that reducing dissonance involves mechanisms influencing the relevance, importance and salience of certain cognitions in order to get rid of problematic ones.[170] Following such a characterisation, it becomes apparent that cognitive dissonance and the way the mind deals with it in order to reduce it seems to tap into some of the mechanisms that manipulative communication aims to tap into. The next section explores this idea in order to assess the role cognitive dissonance theory may play in an account of manipulative and uncooperative communication.

II.2.2.2.2. Discussion: manipulation and dissonance

An important aspect of cognitive dissonance theory should be highlighted and further discussed for the purposes of the present study, namely that an attempt to generate dissonance, for instance by making someone's inconsistencies public, may effectively induce further beliefs or behaviour which the addressee had not planned to adopt and moreover which may run contrary to the addressee's own interest. In particular, and as psychosocial research has shown, speakers can mislead their audience by making apparent inconsistencies relevant in such a way that the audience ends up influenced by something that should in principle not influence them. Take for instance the example of a scholar criticising a theory on the grounds that it shows unable to properly explain a given phenomenon, with a solid argument that objectively makes a point (for instance, imagine the critic presents empirical data which cannot be handled by the theory). Imagine that in order to defend her/his theory, the recipient of the critique counterattacks by highlighting that her/his interlocutor has not come with a better solution to the problem under scrutiny. This in itself is not a defence of the theory, but instead points to the fact that the critic has nothing better to propose. If we formulate the problem in such a way, it becomes apparent that the fact that the critic does not have anything better to offer should not affect the legitimacy of the critique, in particular if the point made is intelligently and intelligibly presented and sustained by the data. However, what the scholar who received the critique is trying to do is to point to some inconsistency of the critic: who is s/he to criticise, when no alternative is offered? This is symptomatic of a variant of the *ad hominem* argument, namely the *circumstantial* variant, in which the speaker, instead of trying to answer the counterargument, files an attack on the counter-arguer on personal inconsistency grounds.[171] Both cognitions involved in the example,

[170] We will see in Chapter III that this constitutes the rationale for a cognitive account of manipulative communication.

[171] Notice that this analysis should also be valid for the *tu quoque* variant of the *ad hominem* fallacy, since it also points as an inconsistency instead of properly addressing the issue at stake.

to take Festinger's terms, are actually irrelevant to each other: whether the attacker is able to provide a better explanation is irrelevant to the validity of the attack s/he filed, because those are two different matters (explaining that B is not a correct explanation of A is different from proposing C as an alternative explanation). And yet, time an again, we find examples of this fallacy that actually work and allow the recipient of the critique to get away with it without having to answer the proper objection.[172]

What this example shows is that pointing out an inconsistency may very well trigger dissonance, and thus induce pressure to reduce it, which can in turn become a hierarchically prominent goal in the exchange (meaning that other operations, such as checking the validity of the message, the trustworthiness of the speaker, etc., become less important or simply do not occur).[173] In the example, the critic feels pressure to resolve the tension between making a point and being portrayed as a somewhat dilettante individual who is deprived of the status of credible interlocutor. What is likely to happen in cases like these is that it becomes more important to the critic to maintain face (in Goffmanian terms) than to stick to the discussion about the counterargument s/he initially provided; it can thus happen that the dissonance between what s/he wants people to think of her/him and what is happening in the exchange, namely that people actually do think of her/him as an unreliable person becomes somewhat inescapable, and that it ought to be dealt with accordingly. Consequently, trying to induce dissonance in the critic turns out to be a way of avoiding a discussion about the original standpoint, namely that the theory under attack is wrong on the grounds that the critic pointed at. It is also important to notice, for further reflection,[174] that this is typical of red herring fallacies, which are meant to distract the course of the exchange from the original standpoint under discussion.

A dramatic illustration of how induction of cognitive dissonance may influence people's beliefs can be found in the work of Edgar Schein, who studied propaganda and indoctrination methods that the Chinese army used on prisoners of war in the Korean War (1950-1953). When American soldiers were captured and deported to prisoner camps, they were neither brutalised nor tortured, but rather welcomed with the words "Congratulations, you've been liberated!" (Schein 1956: 322). A "lenient" policy was enforced: prisoners were not treated as enemies, but as

---

[172] There is probably room to discuss whether such an example is necessarily manipulative; after all, the recipient of the critique is perhaps not deliberately using this strategy in order to avoid answering the objection that was actually brought forth, but simply having a gut reaction which is not consciously designed to mislead the opponent.

[173] We will see further along that dissonance, and the tendency to consistency it exploits, can be assumed to fiddle with the epistemic vigilance filters discussed in Speber *et al*. (forth.)

[174] See §III.5.1.

individuals who had been fooled by the capitalist doctrine, who were thus considered to need enlightenment. The social structure of prisoners was annihilated so that no hierarchical relationships of the sort we find in Western armies were possible; moreover, a suspicious climate was instilled, as prisoners were encouraged to tell on their fellow inmates. However, a collaborative climate was enforced between prisoners and their wardens through dialogue and long discussions. These discussions were meant to allow the Chinese to repeatedly ask from the prisoners that they write autobiographies and that they sign them. As this happened over and over, it constituted valuable opportunities for the wardens to highlight the inconsistencies the prisoners were led to produce in the repetition of the same task. In addition, prisoners were asked to repeat, write down, and sign assertions such as "not everything is perfect in the US", or "there is no unemployment problem in a communist country", that is, assertions the implications of which were apparently not strikingly far-reaching. Once signed, these assertions were broadcast inside the prisoner camp so that everyone could hear what such and such private had said; so whatever they said was bound to be made public, in particular those assertions which were hostile to the US and favourable to the communist doctrine. In turn, these constituted material for the discussion groups that were created subsequently. Little by little, through repetition, the wardens were able to get the prisoners to reflect about ideological matters, but always favourably towards the communist regime, and, more crucially, always knowing that whatever other privates had reportedly said had never been obtained through physical coercion. Everything that was recorded was willingly produced, and this in turn would constitute a reason to consider those messages as actually endorsed by the privates who produced them, especially to the eyes of other prisoners. What the Chinese soldiers were actually doing, as Cialdini (2009) explains, was exploiting the prisoner's natural tendency for consistency. When behaviour is inconsistent with beliefs, usually people try to change the behaviour so that it complies with the belief system. However, in Chinese prisoner camps, behaviour could not be changed, because what the prisoners had said was signed and broadcast to the rest of the camp. In such a case, behaviour cannot be changed, and thus one of the ways of reducing the dissonance is to adapt one's beliefs accordingly. Some prisoners consequently started to gradually modify their initial antipathy towards the communist regime, ending up convinced that the system promoted by the USA was not so perfect after all. This was referred to later on as a clear example of brainwashing techniques. Yet, evidence from the strength and effectiveness of this technique, as pointed out by Schein (1956) is still shaky, to the extent that prisoners having shown sympathy for the communist cause returned to the values promoted by the USA shortly after they were freed and taken back to the American continent.

In any case, this shows that there are grounds to debate whether the mechanisms related to cognitive dissonance so far discussed can indeed be used to characterise manipulative attempts.

In the case of the circumstantial *ad hominem*, the shift from the original standpoint under discussion to the issue about the counter-arguer's management of face in the conversation is actually public (it is in fact what is happening in the exchange and directly noticeable by conversational participants and witnesses), so the only thing that could remain covert are the reasons for which the recipient of the critique is performing such a fallacy. It could be argued that the counter-arguer feels coerced into a defence of her/his face, but even if this is inevitable, it is still something that can be straightforwardly apparent to all. And thus the fallacy itself may considered not to be manipulative after all, since its *modus operandi* is absolutely transparent to anyone keeping track of the exchange, even if the reasons the author of the fallacy has for doing so still may be. The case of the American prisoners of war discussed by Schein (1956) is also debatable as an illustration of manipulation, because the prisoners do have some external motivation for engaging in a behaviour consistent with favourable attitudes towards the communist regime: they can always justify their behaviour by the rewards they would get, such as cigarettes or fruit, and they may even motivate it by fear of retaliation or punishment sanctioning failure to comply with instructions from the wardens.

Now, even if the idea of generating cognitive dissonance in order to manipulate remains a question for debate, there are alternative grounds to postulate that manipulation is precisely about the opposite, namely about preventing emergence of a specific cognitive dissonance between two particular cognitions. As dissonance is defined in theory as an incompatibility between any given two cognitions, we could imagine the case where one of those two cognitions concerns assumptions that the addressee might hold about the speaker's cooperativeness, benevolence and degree of honesty in the exchange, and that the other concerns the belief that the speaker is trying to get the hearer to entertain (let us call it P). Let us focus in particular on these cognitions when a manipulative attempt has been identified by its target, or when the target starts having doubts about the speaker's overall cooperativeness. There are no obvious counter-arguments to oppose to the standpoint according to which the situation in which the hearer spots that the speaker tried to manipulate him into entertaining belief P can be interpreted as a case of cognitive dissonance: on the one hand, there is a cognition about the speaker's deceptive intention, and on the other, a cognition about P. Those two cognitions will presumably become dissonant the moment the manipulative attempt is identified, because usually spotting the manipulative attempt will negatively affect reasons for believing P (since it is recognised that the speaker wants to manipulate the hearer into believing P). In other words, once the hearer comes to know that the speaker has attempted to manipulate him, the likeliness of the hearer adopting P will decrease as a result of this dissonance. Taking things from a different angle, we could thus argue that the success of manipulation, at least on this theoretical level, crucially depends on whether the manipulator is able to prevent the hearer from

experiencing a dissonance between cognitions related to the manipulative attempt and cognitions about the content of the message. Since experiencing dissonance puts pressure on the organism to reduce it and encourages the individual to engage in further processing of one (or both) of the dissonant cognitions, one could argue that this specific dissonance, if experienced, will prompt the hearer to look for a way of getting rid of it, and this would potentially be achieved by further reflection about the speaker's motivation in wanting the hearer to adopt P. This is typically what manipulators want to avoid, because it increases the chances of inferring that something is fishy in the communicative exchange.

To sum up, following cognitive dissonance theory, we could argue that there is one particular cognitive dissonance that manipulation absolutely needs to avoid because it would potentially lead to making the manipulative attempt fail, and it is the dissonance that would be generated between cognitions about the speaker's intentions and cognitions about what the speaker is asking from the hearer, in terms of behaviour or belief adoption.

II.2.3. CONCLUSION

We have seen in this section on social psychology that a significant part of research on influence and persuasion is highly relevant to an informed account of manipulation, to the extent that this field of research has identified constraints on behaviour and cognition that play a crucial role when it comes to influencing people.

However, most of these accounts have limited themselves to identifying and explaining these mechanisms, either from a social or cognitive viewpoint, without specifically trying to assess their linguistic and communicational counterparts. I have already begun to give a few examples of how these mechanisms could be discursively realised by pointing to some striking correspondences in particular between argumentative fallacies and principles of influence. The purpose of chapter III will be to look into more detail at these communicational counterparts of cognitive and social constraints on behaviour and cognition.

Thus, what I will be out for is a general framework which will tentatively be able to integrate the input of social psychology (both sociologically- and psychologically-inspired) but also of cognitive psychology, as we will see next. What I will hence be needing for this purpose is a theoretical model of communication that is able to take into account both external and internal constraints on information-processing, as I will develop the idea that the success of manipulation involves the discursive exploitation of the type of mechanisms that have been discussed in the present section, and that are meant to tap into the latter in order to achieve influence and manipulation through communicative exchanges. Crucially, I will look at

information-processing mechanisms, in particular the interpretation of meaning, so as to see how different constraints designed to influence can be exploited through communication.

There are epistemological implications associated to the naturalistic ontology for mechanistic explanations I am opting for, in Sperber's vein (see Sperber forth.). The ancient debate between sociologically- and psychologically-inspired accounts of human behaviour is obviously at the very centre of the issue, in particular in what regards the study of human communication. Some trends consider that the social sciences cannot be put on a par with the so-called 'hard sciences', by virtue of a purported significance of the difference in nature of the phenomena each of these fields sets out to account for. Sometimes scholars in the social sciences assume that the phenomena that can be addressed, say, by Physics, are completely different from phenomena related to human interaction because the latter seem to be infinitely more complex than the former and that experimental conditions cannot be controlled in the latter case the way they can be in the former; they conclude from this difference that the methods used to study them should be different. Such a conclusion is nonetheless questionable: first, the principles governing gravity, for instance, are themselves infinitely complex and furthermore subject to variation depending on environmental conditions, and thus justifying the existence of competing epistemologies on the grounds of the simplicity of some phenomena and the complexity of others is not sufficient to the extent that simplicity and complexity are hard to measure. Second, just because the phenomena that we want to study are of different nature does not justify that we should use different epistemologies to address them; some methodological principles, such as laws of non-contradiction and deductive reasoning are scientifically strong enough to characterise all scientific endeavours. Third, the robustness and constant replication of the effects observed in the social sciences reveal a regularity that affords to be studied, at least in principle, with the same type of methodology used in hard sciences, which is also targeted at assessing and accounting for the regularity of phenomena. Finally, it should be stressed that there would be no social facts without human minds; in particular, regarding language and communication, as Chilton – when discussing CDA's stance on the issue – puts it, we should accept that

> "if language use (discourse) is, as the tenets of CDA assert, connected to the 'construction' of knowledge about social objects, identities, processes, etc., then that construction can only be taking place in the minds of (interacting) individuals." (Chilton 2005a: 23)

Taking these arguments into consideration, it indeed seems quite legitimate to inquire about the cognitive underpinnings of the phenomena characteristic of social interaction; in the next section I start addressing these issues more in detail as I try to assess the type of cognitive constraints that could play a role in manipulative communication.

II.3. COGNITIVE CONSTRAINTS ON INFORMATION-PROCESSING

II.3.0. INTRODUCTION

As a branch of the broader cognitive science paradigm,[175] the cognitive psychological agenda started off as an attempt to overcome the limits set by the pre-1950's radical behaviourist stance according to which the human mind is an inscrutable 'black box' whose study has little scientific value to offer, leaving observable behaviour as the only possible object of study.[176] In reaction to this methodological principle, cognitive psychology undertook the task of precisely trying to model the type of cognitive operations at work when our minds process information and which are responsible for how we behave. The two major theoretical assumptions constituting the bedrock of cognitive science are in this respect i) that mental activity involves the ability to represent states of affairs in a symbolic fashion and ii) that those representations can be combined with each other in a computational fashion according to procedures which can in turn be described. The most important principle that cognitive psychology draws from these assumptions is that the description of the way these procedures unfold is a first step in the explanation of the phenomena they underlie. As Thagard (2010) summarises it, "the central hypothesis of cognitive science is that thinking can best be understood in terms of representational structures in the mind and computational procedures that operate on those structures". From such a perspective, it is very common to resort to the mind-computer analogy: the mind would be to the brain what software is to hardware. The human brain is thus conceived as some kind of information-processing device which takes perceptual input, encodes it to turn it into representational material which is then processed together with other representations in order to yield an output, in the form of another representation, mental states or actual behaviour. The task of cognitive science is thus targeted to produce psychologically plausible models of these information-processing procedures, and is typically concerned with phenomena such as reasoning, memory, judgement, analogy, concept formation, language competence, etc.

The study of reasoning is particularly worth of attention with respect to the issue of manipulation as many manipulative strategies are meant to mislead their targets by trying to hide problematic articulations between premises which lead to the communicated conclusion (I am thinking in particular of formal fallacies which can deliberately be meant take advantage of their resemblance with formally valid arguments in order to go unnoticed and eventually convince). Studying reasoning requires that we first have a rough idea of what reasoning is, that

[175] Cognitive science is here understood as the science of the human mind, taken as information-processing system.

[176] See for instance Skinner (1953), Watson (1913)

is, that we have at our disposal some standard that we can reasonably take to characterise the reasoning ability. As mentioned earlier (see §II.1.2.2.), since it provided highly reliable and failsafe reasoning patterns (such as deduction), formal logic was traditionally regarded as a reliable representation of human reasoning abilities; consequently, it was also regarded as the most promising reasoning system to be implemented on computers within artificial intelligence programmes. This in turn was accompanied by the idea that formal logic and the criterion of formal validity could be taken to characterise rationality altogether, making irrationality correspond to illogicality.

Yet, the predominant role researchers had been intending to give formal logic in terms of standards of reasoning and rationality became a matter of great debate. With time and psychological experimentation making progress, researchers indeed gradually came to terms with the idea that formal logic alone was not enough to characterise human reasoning (cf. Andler 2004: 316-317, who develops the following four arguments): first of all, artificial intelligence showed the limitations of logic as the latter failed to succeed in the tasks that we once thought it would flawlessly carry out. This is also an idea expressed by Sperber and Wilson: "if there is one conclusion to be drawn from work in artificial intelligence, it is that most cognitive processes are so complex that they must be modelled in terms of heuristics rather than failsafe algorithms" (Sperber & Wilson 1995: 45). Second, the proliferation of alternative logical systems led to questioning the link that was once thought to be exclusive between formal logic and reasoning. Third, connexionist approaches to cognition show that representation and computation-based accounts can be challenged by other models which do not rely on a formal logic component but reason in terms of neuronal network activation. And fourth, flourishing evidence stemming from experimental research steadily started to shake the foundations of formal logic's supremacy, in particular the Wason Selection Task paradigm and the Heuristics and Biases programme initiated by Tversky and Kahneman, as we shall see next.

It is thus in order to test if logical standards were indeed faithful formulations of mental processes involved in reasoning activities that all these studies in cognitive psychology emerged. In what follows I review in more detail than earlier (see §II.1.2.4.2.) the Wason Selection Task paradigm and the Heuristic and biases programme before moving on to highlighting the significance of cognitive biases and errors for an account of manipulative communication.

II.3.1. REASONING, BIASES AND HEURISTICS

In reaction to the logical approach to cognition, an extremely prolific experimental paradigm of research developed in the 1960's, first under the impulsion of Peter Wason and later on with the work of Daniel Kahneman and Amos Tversky, who laid the foundations of the psychological

study of judgement and decision-making under uncertainty (culminating with prospect theory). Both strands of research, albeit in different ways, have considerably challenged the view according to which formal logic is the correct standard for reasoning, leading to a whole rethinking of the nature of rationality. I will first deal with the Wason Selection Task paradigm and then move on to examining the main tenets of the heuristics and biases paradigm, before I conclude the section by highlighting the link between information-processing, communication and reasoning.

II.3.1.1. The Wason Selection Task Paradigm

II.3.1.1.1. Overview and contemporary reinterpretations

I have evoked earlier (§II.1.2.4.2.) the original purpose of the Wason Selection Task (henceforth WST) and also presented its experimental design as I was developing arguments against the suitability of formal logic, but with respect to standards of argument soundness. I will now return to this and try to highlight the implications of the WST in terms of what it tells us about our reasoning skills (in particular conditional reasoning. I will thus summarise what I have already said but also amplify the account, notably in light of Sperber *et al*.'s reinterpretation of the whole paradigm, and of Andler's (2004) comments on the relevance of the WST to the study of human reasoning.

The WST allegedly shows that people perform very poorly at abstract tasks whose success requires that participants perform conditional logical reasoning (i.e. reasoning patterns containing a major premise of the 'if...then' type): numerous experiments and later replications reveal up to 80-90% of 'illogical answers' in these tasks. Let us recall that the gist of the WST was to ask subjects to select an appropriate number of cards – among four cards in front of them, each of which has a letter on one side and a number on the other – allowing them to evaluate whether a conditional rule establishing what is on both sides of the cards is true, typically something like "If there's an A on one side of the card, there's a 7 on the other" (see figure 3).

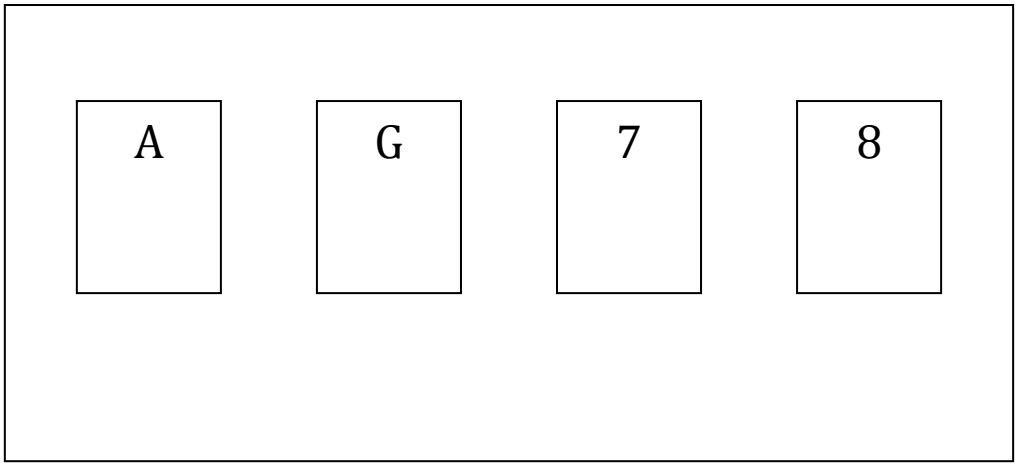


Figure 3. The Wason Selection Task design.

It was straightforwardly assumed from the results that failure to select the cards complying with the logical solution was evidence that people were not spontaneously performing the required logical operations, and this in turn constituted one argument against the supremacy of formal logic as a reliable standard of reasoning: if, when faced with a task whose resolution requires logical reasoning, subjects fail to reason logically, why should we keep formal logic as the actual norm of reasoning? For keeping the standard of formal logic would entail that people are wholly irrational, which raises further problems. Wason explained his results through what he called the confirmation bias, according to which people are more prone to look for, process and store information which confirms what they already believe than to look for falsifying evidence. According to Wason, the task he designed should have prompted instead for a (Popperian) falsificationist strategy:

> "The kind of attitude which this task demands is that implicit in the formal analysis of scientific procedure proposed by Popper (1959). It consists in a willingness to attempt to falsify hypotheses, and thus to test those intuitive ideas which so often carry the feeling of certitude." (Wason 1960: 139).[177]

Yet, the results show that people do not typically opt for disconfirming cards, which Wason interpreted as the influence of the confirmation bias: according to him subjects took the rule as true and only looked for information that would confirm it. This interpretation was threatened by a later experiment by Evans and Lynch (1973), who wondered whether people selected the cards following the formulation of the rule, that is, whether people selected only the cards that

[177] Although this quote is taken from a paper dealing with another of Wason's experiments (the '2-4-6' task), it is safe to say that the confirmation bias was thought by Wason to characterise the Selection Task as well.

were named in the rule, as some kind of label priming mechanism. In order to test this, they modified the rule into a negative formulation ("If there is an A on one side of the card, then there is not a 7 on the other"). If subjects have a confirmation bias, they should go for the A and 8 (or non-7) cards, in order to discover the card about which the rule precisely is. If they have a matching bias, however, they should pick A and 7, since those are the cards explicitly named in the experiment. The results favoured the matching bias hypothesis, but it was later on established that the matching bias could be reduced with some manipulations of implicit negation in the formulation of the rule. This is not something I have time to go into detail here, but let me nonetheless note that it brings forth a very significant observation: changes in the formulation, that is, linguistic/communicative variation, may induce changes in performance.

A few pieces of research contend that the Wason Selection Task shows problematic if we consider that it is aimed at testing logical reasoning. This is an idea that Sperber, Cara and Girotto (1995) among others have defended with convincing arguments. Their treatment of the Wason Selection Task contends that the task does not in fact allow us to test the subject's logical competence. It is not a reasoning task proper, since the subject is not instructed to produce a conclusion which should follow from specific premises: in other words, the subject is not instructed to actually perform *modus ponens* or *modus tollens* reasoning patterns. He is instead instructed to select the cards that he thinks are relevant to resolving the task.

Andler (2004: 359) is also among those who contend that the WST is ill-formulated as a task of logical reasoning: he notes that there is a problem in the formulation of the task to the extent that the material implication associated in formal logic to the 'if…then' form is actually very odd in natural language. If we take the truth table of the implication, the last two lines (where P is false, Q true, the implication being true, and where P is false, Q false, the implication being true) are so unintuitive that one would be hard-pressed to try to make sense of them. If the rule stipulates that each time P is the case, Q is *ipso facto* the case, then why would we care about situations in which P is not the case? Moreover, what can we possibly make of an implication when we know that its antecedent is not the case? There is *a priori* little relevance in reflecting about the implication if we know beforehand that its antecedent is not verified. Adding to this that 'if' is quite often interpreted in natural language as a bi-conditional (iff), such as in "if you wash my car, I'll give you 30 francs" (meaning "if and only if", and that the only scenario in which I will give my interlocutor 30 francs is if he washes my car), it seems that the WST is ill-designed: on the one hand, it aims at testing a logical concept (the material implication) which is in part irrelevant to our natural reasoning needs (Andler does not see in what way the last two lines of the truth-table of the implication are useful to us), and on the other it fails to properly deal with

the difficulty that the logical concept under scrutiny may have a different interpretation in natural language (i.e. the bi-conditional). Unsurprisingly, Andler concludes that:

> "if this 'if' [the logical if] does not exist, what can the subject, whom the experimenter is trying to get to interpret the natural 'if' as a logical 'if', do? We deceive him, we trick him: he is not placed in a normal problem resolution situation. He must face a perverse situation that he has trouble grasping […]." (Andler 2004: 367, my translation)

It appears that asking from the subjects that they understand the logical 'if…then' is pointless, since it would be like expecting them to understand something in a different way than what they are spontaneously inclined to do, and in this sense the WST could be considered as deceitful.

Should the WST have been a proper task of logical reasoning, it would have needed to state premises and then the subject should have been asked about an ensuing conclusion; for instance, s/he should have been given first a conditional rule ('If P then Q'), then the antecedent (P), and then s/he should have been asked to state, given this information, what conclusion could necessarily be drawn. Andler contends that if the task were formulated like that, few people would get it wrong; since such a version would correspond to a proper reasoning task, we would probably have been inclined to conclude that people's reasoning abilities are indeed logical. Thus, the WST, as it has been formulated originally, does not show that we are not effective deductive reasoners. In support of this last idea, it should also be emphasised that once you explain the task to the people, they are usually able to understand it, which testifies in turn of their mastery of a logical ability. Just because they did not do so in the first place is thus not the sign that they are unable to reason logically, but probably rather that there is something in the task's formulation which is responsible for such low success rates in the experiment.

This same idea precisely lies at the core of Sperber *et al*.'s (1995) understanding of the WST. As explained before (§II.1.2.4.2.), they came up with an alternative explanation which does not rest on presumably flawed logical reasoning, but rather on the proper functioning of our comprehension mechanisms, in terms of relevance. The idea is that subjects fail because they are unable to find relevance in the formulation of the abstract task (which is not the case in the deontic version, since the latter actually does raise expectations of relevance for the correct solution); accordingly, they ventured that a way of boosting performance in the task is to make the expected answer's relevance high in the context. For this, Sperber *et al*. came up with a "recipe" for generating easily-solvable selection tasks: fist, one should "select a pair of simple features *P-and-Q* such that the complex feature *P-and-(-not-Q)* is, or can be made, easier to represent than the complex feature *P-and-Q*" (Sperber *et al*. 1995: 60). The second step is to "create a context where knowing whether there are *P-and-(not-Q)* cases would have greater

cognitive effects than knowing whether there are *P-and-Q* cases" (*ibid.*). The third step is to "present the rule 'if P, then Q', in a pragmatically felicitous manner" (*ibid.*).

To illustrate how this recipe works, let us report one of the experiments Sperber *et al*. carried out. Subjects were presented with a text narrating the story of a journalist's investigation about a cult whose leader was suspected of artificially inseminating virgin girls in order to create some kind of elite class of 'pure' mothers. The leader of the cult had categorically denied such allegations with a conditional statement such as "if a woman has a child, then she has had sex" (and this is the conditional rule that the subjects will have to evaluate in the task). The narration then reports that the reporter paid a visit to the gynaecologist who had examined all the women of the cult. Invoking medical secrecy, the doctor refuses to answer the reporter's questions, but the latter notices four half-covered information cards on the table between the doctor and him; one of them reads "Children: yes", the second "Children: no", the third "Sex: yes", and the last one "Sex: no", but the reporter can only see the top half or the bottom half of each card. By uncovering the other half of these reports, the reporter could learn the missing information (whether the women have children and whether they have had sex) that would reveal if the guru was lying or not. Figure 4. shows the material subjects were given:

Until recently, it was obvious that a woman who has children has had sex. With artificial insemination, it is now possible for a virgin to have children. The leader of the Haré Mantra (a very secret religious, Californian sect) has been accused of having had some of his sect's virgin girls artificially inseminated.

His goal, it is claimed, is to create an elite of "Virgin-Mothers" alleged to play a particular role in his religion. The head of the Haré Mantra makes a joke out of these allegations. He claims that the women of his sect are, without exception, like any other women : **if a woman has a child, she has had sex**.

Imagine that you are a journalist and that you are preparing an article on the Haré Mantra. You learn that a gynecological survey has been carried out among the Haré Mantra women. Some of them might be "Virgin Mothers". You go and visit the doctor who carried out the gynecological survey. Unfortunately, the doctor pleads professional secrecy and refuses to tell you what he discovered.

You realise that, before you on the doctor's desk, there are four individual information cards about Haré Mantra women examined in the gynecological survey. However, these four cards are partially concealed by other papers (as shown below). Of two cards, one can only see the top where you can read whether the woman has children or not. Of the other two cards, you can only see the bottom where you see whether the woman has had sex or not. You are determined to take advantage of a moment in which the doctor turns his back to uncover the papers and to learn more .

Indicate (by circling them) those cards that you should uncover in order to find out whether what the leader of the Haré Mantra says (**"if a woman has a child, she has had sex")** is true, as far as these four women are concerned. Indicate only those cards that it would be absolutely necessary to uncover.

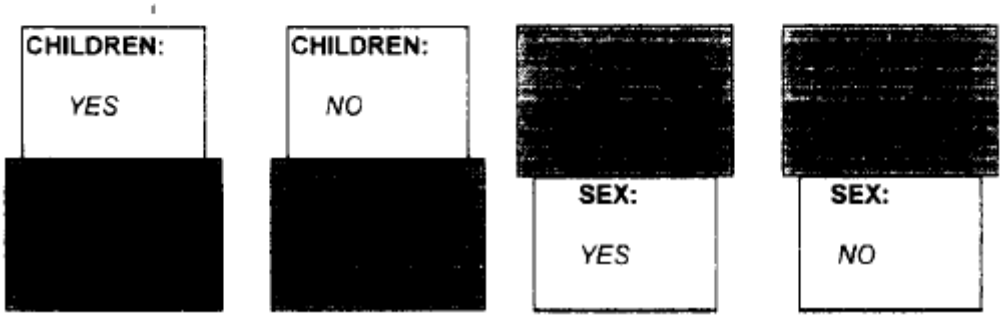


Figure 4. 'Easy' version of the selection task (Sperber *et al.*1995: 63)

The first step of the recipe for easily-solvable selection tasks stipulates that the correct answer, *P-and-(not-Q)*, has to be made relevant, and this is achieved in the narration by making the story revolve around the discovery of *P-and-(not-Q)* items – i.e., items combining the features 'mother' and 'has not had sex' –, instead of looking for *P-and-Q* items – i.e., items combining the features 'mother' and 'has had sex'. The virgin-motherhood combination is so salient in the text that the natural associations, 'virgin-non-motherhood' and 'motherhood-non-virgin', are less likely to be represented as relevant options. The second step of the recipe is fulfilled as the correct combination ('virgin-mother') is particularly relevant in the context: for there to be a successful news story, the reporter needs to discover *P-and-(not-Q)* items, since this is the combination which is likely to yield rich implications for the reporter. The third step consisting in coming up with a pragmatically felicitous formulation of the rule is fulfilled by putting it in the mouth of the

cult leader as a defensive statement, which prompts the reporter to want to falsify it, instead of confirming it (which is what happened in the original WST). In addition, the leader uses a conditional statement instead of a clear-cut denial, which functions as an appeal to the audience's common sense dictating that usually women with children have had sex. The results of this version of the task are striking, as 78% of the subjects succeeded in picking the correct cards.

II.3.1.1.2. Discussion

Summing up, if we go by Sperber *et al.*'s (1995) and Andler's (2004) understanding of the WST, it should be clear by now that such a task is not suited for testing logical reasoning. Instead, it seems that this type of experimentation has quite a lot to tell us about the importance of verbal comprehension, in terms of information-processing, for reasoning. If subjects fail to produce a correct answer, it is not because they are poor logical reasoners, but because the way the task is formulated does not encourage them to engage in logical reasoning and ends up orienting them on the wrong track. Under this view, differences in wording and in the generation of expectations of relevance can be actually said to mislead subjects, so much so that tweaking those parameters results in differences in performance. This in itself is significant to an account of manipulative communication because it precisely shows that people's cognitive operations can be constrained by manipulating expectations of relevance. To take things a step further, as Andler (2004) does, there is even room to consider that the WST, in its original wording, is actually misleading in that it fools subjects into processing the information in a way which will make it really hard for them to discover the correct answer.

The WST is not to be taken as an empirical argument against formal logic. To take Andler's analogy with language, it is not because we make logical errors that a logical competence does not exist (just like it is not because we make linguistic errors in performance that there is no language competence altogether). Rather, what the selection task shows is that one can prompt for poor logical reasoning performance if certain conditions are met, among which the generation of particular expectations of relevance which turn out to be particularly incompatible with logical reasoning, and this is something that happens at the level of basic information-processing mechanisms of verbal comprehension.

The output of the comprehension of communicated material (which is driven by considerations of relevance) provides the input for further reasoning about the material that has been communicated. We have seen with this review of the WST that the way relevance will be achieved is subject to biases and errors; thus, whatever output comprehension yields runs the risk of compromising further reasoning. This means that comprehension of communicated

material may already affect belief fixation prior to any processing typically associated with reasoning or argumentation, to the extent that understanding what the task is about should be achieved prior to solving to task. In other words, a representation about knowing what to do functions as an input representation (presumably a belief) for the reasoning task itself. To think of it, this should not be surprising, as it rests on a linguistic truism: depending on what we understand from a given utterance, ensuing beliefs and behaviours may vary. An example of misunderstanding given by Timothy Chambers, in the Korea Times (October 13, 2005) illustrates this simple point: "I told the manager about a problem I was having with the audio-visual equipment, and he wouldn't help me at all! He just apathetically asked, 'What do you want me to do about it?'". The author explains that he interpreted at first this last sentence as an implicit means of manifesting lack of concern on behalf of the speaker; however, he found out afterwards that the speaker really meant, literally, to ask what he could do to help. Depending on the way the input is interpreted and thus depending on the intentions conversational participants are able to attribute each other, different reactions will ensue: in this case the hearer will form a negative appreciation of the manager's willingness to help, which could then lead to the belief that the manager is generally unhelpful. If the pragmatic misunderstanding is not resolved, such a belief could well stay in the hearer's mind and be stored as a reliable representation.

Even if this may sound trivial, it should nevertheless be emphasised that whatever people interpret from communicative material may already play a role in belief fixation. To the extent that the success of communication cannot be guaranteed by a failsafe algorithm (cf. Sperber & Wilson 1995), this is of particular importance for the issue of manipulative communication: I will argue in this respect that the interpretative mechanisms responsible for the generation of misunderstandings are the same that are responsible for successful manipulation. That is, the inherent fallibility of interpretative mechanisms is also responsible for the success of manipulation.

Where does that leave us? To sum up the implications of the WST as far as the cognitive underpinnings of reasoning are concerned, let me state the following:

- The WST paradigm does not prove that people have poor logical skills.
- The WST paradigm experimentally confirms that the mind is a fallible information-processing device when it comes to logical reasoning.
- The WST favours the assumption that certain heuristic procedures and cognitive biases may be at play in information-processing mechanisms (which could be the result of time, available resources and efficiency constraints).

- Comprehension procedures (which are typically of the heuristic kind as they are automatic, fallible and spontaneous) provide inputs for reasoning processes, the outputs of which are mental states or behaviour.
- Thus, the study of verbal comprehension and of its cognitive underpinnings is relevant to the study of reasoning errors to explain the success of fallacies, which is also a promising contribution to the study of manipulative communication.

II.3.1.2. Kahneman and Tversky's heuristic and biases programme

II.3.1.2.1. Overview

Daniel Kahneman and Amos Tversky are the pioneers of a strand of research in cognitive psychology which study decision-making and judgement under uncertainty in order to understand how people make decisions when they lack sufficient knowledge, time, or resources to make an informed choice. In the 1970's, they began to carry out experimental work that specifically tested the way people reason about probabilities in different settings involving various aspects of probability theory. We owe them famous experiments such as the Linda Problem (see Tversky & Kahneman 1983 but also Fisk 2004 for a comprehensive review of the literature on this experiment and on the conjunction fallacy more broadly) or experiments on Bayesian inference (see Kahneman & Tversky 1973 and Birnbaum 2004 for a review) that bring into light several violations of probabilistic theory which nonetheless tend to characterise the way we spontaneously reason about probabilities. Kahneman and Tversky studied in this respect specific heuristics, such as the availability heuristic, according to which people estimate frequencies of occurrence depending on the availability of items,[178] and the representativeness heuristic, which allows people to quickly decide the probability of a given outcome based on its similarity with its 'parent population' (see Kahneman & Tversky 1972). These experiments all point to a gap between the theory of probabilities that has been elaborated by logicians and mathematicians and people's performance in probability-related tasks, thereby constituting a major source of counter-arguments for the logical view on cognition.

More importantly, following the results found within this paradigm, it appears quite clearly that the mind is subject to cognitive biases. It should first be stressed that these biases usually perform quite well given that they are effort-sparing mechanisms (in terms of the cognitive resources allotted to the operations) which allow most of the times accurate fast processing of

[178] This in turn has been used in stereotype research, as it predicts that someone witnessing high occurrence of individuals with property P that belong to group G will associate property P to all members of G (see Reber 2004: 152-153 for a review).

information. This is perhaps why they have survived evolution: they allow us to provide satisfactory answers while at the same time requiring few resources. However, there is a trade-off to these biases and heuristics since they are not perfect: Kahneman and Tversky's have shown that these are fallible procedures and that they may lead to so-called cognitive illusions (see Pohl 2004). That is, they may generate a perception, a judgement or a memory "that reliably deviates from 'reality'" (Pohl 2004: 2).[179]

II.3.1.2.2. Discussion

Kahneman and Tversky's theoretical conclusions on heuristics and biases have been challenged in the past by other cognitive and evolutionary psychologists (among which Gerd Gigerenzer, Leda Cosmides and John Tooby) who doubted the overall adequacy and relevance of the notion of cognitive bias to the study of mental processes. In order to support this idea, it was shown that since cognitive biases may disappear when certain parameters are tweaked, they are not stable, and thus there is no reason to consider that they are characteristic of human rationality. Cosmides & Tooby (1996) gave experimental evidence in favour of the idea that people were actually good at solving the type of tasks studied by Kahneman & Tversky if instead of framing them in terms of probabilities they were framed in terms of frequencies, which made the conjunction fallacy taken to explain the Linda problem simply disappear (cf. Cosmides & Tooby 1996). Gigerenzer (1991, 1996) also developed arguments to show that the heuristic and biases programme was mistaken on several points:[180] first, the whole paradigm presents a one-sided perspective, inasmuch as it tends to focus only on 'derailments' for which heuristics are responsible, instead of devoting at least equal attention to cases were these prove advantageous; second, the biases identified by research could be taken as artificial constructs, since they can be reduced, inverted, or even disappear depending both on experimental conditions and on theoretical definitions; third, Gigerenzer argues that the whole heuristic programme lacks explanatory power, given that according to him it produces more descriptions than actually testable theoretical models and fails to explain why it is that these heuristics kick in at a given point in information-processing mechanisms.

---

[179] In this respect, it could be claimed that this is particularly vividly illustrated by misunderstandings in communication, which suggests that comprehension procedures are heuristic in nature. See Allott 2007 for a development of this very idea within a Relevance-theoretic framework, and also §II.3.3. and §III.2.1. for a development.

[180] See Gigerenzer (1991, 1996) for the details of these counter-arguments, but also Pohl (2004: 10-14) for a brief yet informative summary and discussion of Gigerenzer's objections.

Despite criticism some striking facts about the way people process information do persist and that these do point to cognitive specificities which are responsible for some errors which are extremely relevant to the study of manipulation. In what follows I will specifically address the communicative dimension of some well-known biases that I think may play a decisive role in manipulative communication.

II.3.1.3. Biases induced through communication

Literature on biases and heuristics in cognitive psychology has seldom chosen to consistently explore their communicative aspects;[181] yet, I have already pointed out, notably through Sperber *et al*.'s (1995) reinterpretation of the WST, that insights into comprehension procedures – and thus linguistic and pragmatic aspects of information-processing – can be instrumental to a proper explanation of cognitive biases. From a broader perspective, it should even be acknowledged that insofar as these cognitive biases are studied experimentally, and thus somehow always elicited through interaction, their occurrence is systematically examined within communicative situations where some input is provided to the subjects in order to study their performance in different tasks. This should be strong enough a reason to add to the study of cognitive biases a layer concerned with communication, as the way instructions are processed will provide input for the proper task to be performed.[182]

II.3.1.3.1. Concept retrieval: the Moses illusion

The Moses illusion is a striking example of failure to identify blatant conceptual inadequacy and was first studied by Erickson and Mattson (1981), who asked their subjects "How many animals of each kind did Moses take on the Ark". Surprisingly, they found that more than 80% of tested subjects answered without noticing the anomalous term in the question (it was Noah who saved the animals in the biblical episode of the flood, not Moses); should they have processed the information any further, they probably would have noticed the anomaly, which is the reason Barton & Sanford (1993) have hypothesised that the illusion was due to shallow semantic processing of the concept Moses.

---

[181] Exceptions are Loftus' work on memory and how memory can be played with (see e.g. Loftus 1975, 1993, Loftus & Palmer 1974) and Pohl's latest work on the notion of 'suggestionality', which tries to capture the communicational aspects of cognitive biases (cf. Pohl 2010). However, since research on this topic is only beginning and has not sought collaboration with linguists and communication specialists so far, there is not much that can be said at the moment.

[182] Again, this is precisely the assumption on which Sperber *et al*.'s (1995) reinterpretation of the WST rests.

In the original (1981) experiment, many variables were tested by Erickson and Mattson in order to identify the conditions of the illusion's success. Through a first experiment, they were able to discard the hypothesis according to which subjects do not actually process the anomalous item by making them reading it out loud (and thus have it linguistically encoded), and also showed that even when subjects were told that some of the material may contain anomalies,[183] the illusion would still work. A second experiment tried to test whether the illusion was due to the question's focus on the animals, by introducing the material as an assertion to be judged as true or false by the subjects ("Moses took two animals of each kind on the Ark"). They found no significant difference, but there is still room to wonder whether the form (assertive or interrogative) changes anything at all, to the extent that both forms require something from the subject: the latter still needs to provide an appropriate response to the task, and this, I contend, is crucial to the success of the illusion, since it concentrates the subject's attention on the task and away from checking conceptual adequacy. A third experiment tested whether the phonological and semantic properties of the two items were determining factors, since Noah and Moses are similar in sound (nasal initial sound, long 'o', two syllables, the first of which is stressed) and in conceptual content (both are main figures in their respective biblical stories, both received messages from God, both have something to do with water, both led their followers out of adversity, etc.). Results show that phonological similarity is not enough (the illusion remains even when Moses is replaced by Adam), but that semantic similarity between the two terms does play a role, which leads Erickson and Mattson to conclude that "the fewer semantic features of a word that fit into the description (i.e., in these experiments, the more dissimilar the inconsistent word is to the correct word), the likelier it is that the inconsistency will be noticed" (Erickson & Mattson 1981: 550-551).

Barton & Sanford (1993) took the explanation a step further by postulating that an additional determining factor of this illusion's success is the extent to which the anomalous item's shows global good fit to the scenario, or context, in which it is used: "if there is a good match on the basis of a few features, further processing may not necessarily take place" (Barton & Sanford 1993: 478). This suggests that the more the anomalous item seems to fit the context, the less further checking will be carried out and the more semantic shallow processing will be favoured.

According to Allott & Rubio Fernández (2002), these are not semantic illusions but pragmatic ones, because they assume that changes in the non-explicitly conveyed content can make a difference in subjects' performance. That is, when the semantic relatedness between the

[183] They were given an example such as "Why was President Gerald Ford forced to resign?" (Ford was never forced to resign) and they were instructed to answer "wrong" whenever they encountered cases such as this one. So they knew what to do when confronted to anomalies.

anomalous item and the correct one is left intact but the pragmatic meaning is modified, changes in performance are accordingly observed.[184] In particular, they claim that this is a misleading instance of *ad hoc* concept formation (cf. Carston 1996, 2002a, 2002b, Sperber & Wilson 1997, and also §III.2.2.): the concept Moses is not accessed in its entirety, and only a limited subset of its properties (some but not all logical and encyclopaedic entries of the concept stored in long-term memory) is mobilised in the process of understanding the point of the utterance. This means that the subject will probably not mobilise assumptions that would be incompatible with what the utterance expresses otherwise literally (for example an encyclopaedic entry that states that Moses is the one who led the people of Israel out of Egypt, which would then logically yield a contradiction with the information that he took animals on the Ark).

The Moses illusion is consequently described in terms of the specific type of pragmatic processing it involves. But this is not sufficient if we want to understand why the concept is shallow-processed; to make for this shortcoming, Allot and Rubio-Fernández hypothesise that the context in which the problematic utterance is produced plays a decisive role: we are in an experimental setting, and a very specific task is expected from the subjects, in that they have either to answer a question or judge the validity of a claim, which are specific goals for which the subjects need to allocate cognitive resources. So the main goal of the task, which is the one towards which resources will be allocated in priority, is not to perform lexical consistency checking, but indeed to answer a question. I claim that this a way of contextually constraining the way subjects will process the utterance. As noted by Allott and Rubio-Fernández, "relevance can be attained by processing only to a level where the subject can provide an answer, rather than the deeper level required to evaluate the question for internal consistency" (2002: 99). By this, they mean that the relevance of processing the information lies in providing an appropriate answer, not in checking whether the concepts used in the utterance are indeed the correct ones, which can also be taken to follow from general conversational conventions such as Grice's CP, or more generally about people's expectations about communication in general. In a nutshell, subjects are misled because the nature of the task drives them away from finding relevance in the verification of the appropriateness of the concepts contained in the utterance. This is also probably why even when people are warned beforehand that the utterance may contain an anomaly, the illusion still shows: first, they do not know what kind of anomaly to expect, and second, the contextual constraint of providing an appropriate answer or judgment, which characterises the nature of the task, is still present, and chances are that the fulfilment of this goal will request most of the cognitive resources available at that point.

[184] For the details of the argument, see Allott & Rubio Fernández 2002: 98-101.

Significant implications can be drawn from this type of experiment as far as an account of manipulative communication is concerned, because it brings to light a very particular mechanism of conceptual accommodation which can be deliberately exploited in order to mislead; the Moses illusion, which has been tested with numerous alternative utterances,[185] gathers experimental evidence for the fact that sometimes we shallow-process conceptual entries. To the extent that the goal of these experiments is to mislead subjects so that they fail to notice an anomaly which is covertly and deliberately put in an utterance, it is 'straightforwardly' manipulative, according to the working definition of manipulation I coined in chapter III (see §I.3.2.). As the perlocutionary goal pursued by the experimenter is about getting the subject not to spot the anomaly, I take it, following Allott, that the type of mechanism involved in the Moses illusion can be exploited in manipulative communication strategies; this is something that Allott has shown about the concept 'democracy', which is sometimes meant to be shallow-processed by the hearer so as to only convey positive connotations instead of the full-blow concept which refers to a particular political system (see Allott 2005).

II.3.1.3.2. The impact of words on memory: misinformation effect and labelling effect

A category of specific biases operate on memory; such is the case of the so-called 'misinformation effect', which has been extensively studied in the pioneering work of Elizabeth Loftus (e.g. 1975, 1993). Experimental data shows that people are prone to making errors in retrieving events from memory when presented with information that conflicts with the event that actually take place. Pickrell *et al*. summarise this by the "tendency for post-event misleading information to reduce one's memory accuracy of the original event" (Pickrell, Bernstein & Loftus 2004: 345).

Several experimental designs were tested, and usually they would require the subjects to watch a short movie or a series of slides and then to answer a few questions about what they had seen. Somewhere in the questionnaire was the target utterance, which was the one that would trigger the misinformation effect. For instance, subjects watched a sequence of chronologically-ordered slides where they would see a red car on a road, the same car stopping at a stop sign, turning right and then hitting a pedestrian. The question subsequently asked was something like "did another car pass the red car while it was stopped at the intersection with the *yield/stop* sign?" (the control group was given information consistent with what the slides depicted, i.e. a stop sign, while the group that was being tested had the question with conflicting information, i.e. the

[185] See the appendix of Park & Reder (1991: 403-405) for an exhaustive list of examples testing the Moses illusion effect.

yield sign). The results showed that conflicting information does have an incidence, to the extent that when they were subsequently interrogated about their memories of the slides, many of the misled participants replied that they had indeed seen a yield sign rather than what was in fact on the slides. According to Pickrell *et al*., "these studies demonstrate that misleading post-event information affects what people erroneously report about the past.

Alarming as they may be,[186] these experiments, which have been multiplied, replicated and refined over the years, show that memory is not to be conceived as a box in which you store things and expect them to be the same the next time you are looking for them. Memories change over time, and one of Loftus' main results is that with time, changes can be profound. An additional and very striking observation that can be made on the basis of several experimental designs is that some linguistic mechanisms may prompt for such memory changes. Presupposition, for instance, has been shown to play a role in such mechanisms: in Loftus *et al*.'s (1978) report, for instance, the target question contains the definite article ("the stop sign vs. the yield sign"), which existentially presupposes that there was a stop/yield sign on the slides. Since presuppositions introduce information which is somewhat peripheral to the main point of the utterance, but which is nonetheless necessary for the utterance to make sense (it would make no sense for me to utter "my sister has two children" if I did not in fact have a sister), they offer chances to sneak in information which should not be accepted as common ground. In this respect, this is precisely what happens in the red car example: the successfully misled addressee, driven by standard conversational expectations, will not remember that there was no yield sign, contrary to what the utterance s/he is processing conveys, because s/he treats as given that the presupposed information is correct. As Greco notes, "since what is taken for granted at a certain point in a conversation is not at issue at that moment, it cannot be checked and discussed, if the conversation is proceeding normally. Therefore, being somehow out of our control, what is taken for granted could be false" (Greco 2003: 218-219).

Presuppositional accommodation, the process by which information which was unknown before is quickly and automatically updated in conversation without being the focus of the utterance and whose truth goes unquestioned, can in this respect also be responsible for the generation of false memories. As far as the study of manipulative communication is concerned, this is of crucial importance: suffice it to point out that the fallacy of many questions, also known by its Latin name *plurium interrogationum*, precisely exploits this mechanism. It consists in producing an utterance loaded with question-begging presuppositions, such as the canonical "Have you

[186] In that it greatly challenges the reliability of memory, which is particularly worrying for the legal world, where witnesses are called to testify about events they are asked to remember.

stopped beating your wife?". If the addressee answers the question, this means that he has accommodated the presupposition that he has been beating his wife in the past. The only way to get out of the trap would hence be to answer back something like "I have never beaten my wife", which points to the falsity of the presupposition. Even if the examples used to discuss this fallacy are in principle very clear, and thus very extreme and unlikely to be found in natural conversation, this does not mean that the fallacy may never work: distraction are always possible, and the content of the presupposed information might be very subtle so as not to generate a processing problem whatsoever.

An additional effect that research on memory has identified, which also depends on the wording of communicated information, is the so-called labelling effect (Loftus & Palmer (1974). In an experiment, subjects were presented with a short video clip of a car crash, and later on asked to evaluate the speed at which the cars were going when the accident took place. Estimates depended on whether the verb used to denote the accident was "hit", "bump", "collide with" and "smash into", with the highest estimates obtained under the "smash into" condition. What is more, a week later, the same subjects were questioned about the video they had seen, and experimenters would ask them if they remembered seeing broken glass on the road; those who had been tested under the "smash into" condition were more inclined to answer yes, when in fact there was no glass in the original video. This means that labelling an event in a certain way may have significant impact on the long-term memory an individual will form about that event. Again, in terms of manipulative communication, this turns out to constitute an important finding; I would venture, but this is very speculative, that phenomena such as euphemisms and negative or positive representation of discursive actors, for instance, are precisely geared towards influencing the memory storage of a specific representation or description of an event. In this respect, I believe this kind of reflection about tentative interfaces opens up a promising direction of research; research in cognitive psychology on memory biases coupled with more traditional discourse-analytical approaches to linguistic phenomena of this kind could provide the communicative counterpart to cognitive processes that is missing in discourse-analytical accounts to elaborate psychologically explanatory models of how language may influence, sometimes in manipulative settings, memory and thoughts.

II.3.1.3.3. Epistemic biases: validity effect and mere exposure effect

Biases that are characteristic of people's judgement abilities have been thoroughly studied by cognitive psychologists and bear important implications for the study of argumentation, persuasion and manipulation, given that they may influence our epistemic attitude towards representations. These are biases that will affect the validity that people ascribe to certain

representations and in this sense they constitute valuable illustrations of some of the mechanisms underlying belief fixation, which is typically the outcome sought by the type of communicative situations the present thesis is concerned with.

The validity effect is defined as the enhancement of the perceived validity of a statement that has been repeated, whereby "if information has been heard previously, people are likely to ascribe more truth or validity to it than if they are hearing it for the first time" (Hackett Renner 2004: 201). What is striking is that this tendency has been experimentally observed independently of the initial attitude towards the statement: if the subject was already leaning towards considering the statement as valid, repetition would reinforce her/his belief in the validity of the statement; if the subject initially considered the statement as doubtful, repetition would lower her/his doubts and turn them into favourable attitudes towards the statement's validity.

The validity effect is a thorough description of the widespread belief that public opinion holds about political propaganda, advertising and marketing strategies: the more people are exposed to a message, the more they will believe it is true. However, research in cognitive psychology has not reached consensus on a definite explanation. Many parameters have been under study, but most experiments end up confirming the existence of the bias rather than providing a solid explanation. As Hackett Renner notes, "it is time to move beyond demonstrations that the validity effect occurs, to demonstrations of its extended impact" (Hackett Renner 2004: 211). One thing that can be quite safely said, though, in terms of information-processing, is that repetition of a message makes the message salient (compared to other information that is not mobilised over and over again), in terms of both the familiarity with which the information contained in the message is endowed and the validity – or strength – associated to the message's content. and also increases familiarity with the information contained in the message. These two properties, strength and familiarity (which contribute to making the stimulus salient), are particularly important for a cognitive account of information-processing because they play a role in the amount of cognitive resources one needs to spend in order to mobilise the assumptions associated to the message that has been repeated in the past and also in making sure that the associated assumptions enjoy a privileged epistemic status (as they are valid). Building on the idea that repeating a message makes its content readily available – as one is used to processing it over and over – and increases its perceived validity, I will defend in chapter III the idea that manipulative strategies, insofar as they try to privilege the representation of certain sets of assumptions over others, can exploit such a tendency. Regarding the technical exploitation of the mechanism, Maillat (2010) argued that some fallacies, such as the *ad populum* (see §II.2.1.2.) take advantage of the fact that we have previously experienced repetition of a message by imputing it to others (and usually a significant amount of people); that is, the *ad populum* does

not repeat the message itself, but instantiates repetition by simultaneously attributing the message to a large amount of individuals. This in itself is of value for a cognitive account of persuasiveness, and, in case it is deliberately exploited to elicit judgements of validity about a message that is at best controversial, also to an account of manipulation that claims cognitive grounding. I will return to this in chapter III (see §III.4.1.2.2.).

The mere exposure effect (see Zajonc 1968 for the seminal account) works similarly to the validity effect, in that it is a bias arising out of repeated exposition to the same stimulus, only that it generates judgements of liking instead of validity. The more people will be exposed to a stimulus, the more they will like it. This is something that has been tested on a wide range of stimuli, such as music, paintings, human appearance, objects, words and so on, and helps explaining behaviour (such as voting for example) and attitudes towards people and objects (e.g. aesthetic taste for art, sympathy for people…). The idea that people will have a tendency to like whatever sounds or looks familiar to them is obviously something that resonates with Cialdini's principle of liking, according to which people "prefer to say yes to individuals they know and like" (Cialdini 2009: 172). According to Bornstein & Craver-Lemley (2004), roughly two kinds of accounts can be distinguished nowadays: on the one hand, some advocate an affect-based explanatory model that leaves practically no room for cognitive processing beyond the encoding processes of what is perceived. This is Zajonc's stance, who uses the evidence he gathered over the years to argue that other species, less evolved than mankind, are also prone to experiencing the mere exposure effect, and thus that the latter does not depend on complex cognitive operations. On the other hand, there are those who privilege a cognitive explanation involving representational steps. However, consensus has not been met, and Bornstein and Craver-Lemley venture that given that available evidence actually supports both views, the appropriate answer might lie in a combination of both accounts:

> "It may be that MEEs [Mere Exposure Effects] occur in stages, the first of which is a 'pure' affective response that requires minimal cognitive processing beyond rudimentary encoding of stimulus properties. This initial affective response is then moderated by more extensive cognitive processing of the mental representation of the merely exposed stimulus."[187] (Bornstein & Craver-Lemley 2004: 230)

As a bias influencing attitude towards an object or person, the mere exposure effect is obviously of great interest for speakers with the intention of influencing others. The bias will not affect the

[187] I cannot help but thinking about a parallel to be drawn here between this assumption and dual-process theories (see II.2.3.1.); the affective response's automaticity and intuitiveness seem compatible with heuristic-driven processing while the 'extensive cognitive processing' appears to refer to central or systematic processing of information.

epistemic attitude towards its object *per se*, contrary to the validity effect, but it does operate in such a way that its object is preferred in some way to alternative ones, thus satisfying some conditions that will try to make sure that representation related to the object are part of whatever cognitive processing the subject will go through. This in itself is not manipulative in any way; however, it can become manipulative the moment the speaker trying to induce the mere exposure effect is doing so in order to focus the audience's attention on something in particular, *with the consequence that other representations are not focused on* (typically assumptions that would jeopardise the manipulative attempt altogether). In other words, exploiting the biases presented in this section may serve as a manipulative strategy meant to get the audience not to process alternative sets of assumptions, and in this respect it is of high relevance to the cognitive account of manipulative communication I am after.

II.3.1.3.4. Conclusion: biases, heuristics and manipulative communication

Having reviewed some aspects of cognitive psychological research on biases and heuristics, I am now in position to assess how exactly they may come into play in the account of manipulative communication I will elaborate in the next chapter.

First of all, experimental evidence for cognitive biases and heuristics is compelling in support of the hypothesis that the mind is a fallible information-processing device, since it establishes factual and thus objective warrants for it.[188] Second, the type of research exposed in what precedes helps understanding under which conditions biases occur, and this is crucial for an attempt to trace the circumstances, in particular the communicative circumstances, in which biases are likely to appear. Third, once these conditions are known, we can start looking at the type of linguistic and pragmatic phenomena that seek to exploit the requirements of cognitive biases.

II.3.2. FILTERING INFORMATION: THE CASE FOR EPISTEMIC VIGILANCE, COGNITIVE AND EMOTIONAL FILTERS

II.3.2.1. Epistemic vigilance

Communication is a source of knowledge: one of the ways of knowing more about the world around us and of getting along in it is to listen to what other people have to say about it. While communication evolved presumably so as to promote a reliable way of upgrading people's

[188] I should perhaps also mention here that most of the experiments reported in this chapter have abundantly been replicated on populations varying in age, ethnic origin, socio-cultural environments, gender, etc., and therefore that these studies can reasonably claim universal generalisation.

knowledge about the world in order to maximise chances of survival (cf. Sperber 2000a, see also §I.1.5.), it should also be acknowledged that reliance on communication supposes the risk of being accidentally misled or intentionally deceived. Even if most of the time people do not try to mislead others, the properties of communication are such that manipulation can occur. Cheaters are well-aware (although probably only intuitively for the most part) of the possibilities offered by our imperfect communicative practices; some evolutionary theories precisely claim that communication emerged as an instrument of deception: we have seen earlier that such an assumption underlies Byrne and Whiten's account of macchiavellian intelligence (see §I.1.5.2.). Be that as it may, for communication to end up selected as a stable property of human life supposes that its benefits have outweighed its disadvantages in the long run, and thus that the human species has found a way of dealing with cheaters and deceivers while still preserving communication. For Sperber *et al.*, this presumably means that we have developed some abilities to spot cheaters and to assess the reliability of communicated material, some kind of "suite of cognitive mechanisms for epistemic vigilance, targeted at the risk of being misinformed by others" (Sperber *et al*. 2010: 359). Within this perspective, epistemic vigilance is defined as a cognitive ability which is intimately linked to the possibility of communication, in that it appears to be a necessary condition for communication to stabilise. Sperber *et al*. note: "a disposition to be vigilant is likely to have evolved biologically alongside the ability to communicate in the way that humans do" (Sperber *et al*. 2010: 360-361).

The notion of trust is central to the account. Without trust, communication could not have possibly evolved, since in that case what would have been favoured is systematic scepticism, and though we can call into question what others tell us, this does not appear to be our default attitude in communication. Besides, it would be cognitively extremely costly and inefficient compared to a trusting strategy, because it would leave perception and inference as the only two sources of knowledge. Evidence in support of the idea that trust is presupposed in every communicative exchange is found in philosophical accounts (e.g. Reid 1764, Lewis 1969, Davidson 1984, Grice 1989), but also stems from psychological accounts (Gilbert *et al*. 1990, Gilbert *et al.* 1993), who argue in favour of default human gullibility. However, these accounts are compatible with a principle of epistemic vigilance, as long as epistemic vigilance is conceived as a set of mechanisms that kicks in when circumstances prompt for it. That is, humans are generally trustful until or unless they are able to identify specific cues indicating that they should adopt a more cautious and critical attitude towards the speaker and what the latter is saying. This means that whenever one is confronted to communicated information, two distinct processes will come into play: one process will try to make sense of the message (i.e. to establish its contextual relevance so as to achieve comprehension), and the other will assess whether the message is trustworthy. From a theoretical perspective, this establishes a very intimate link

between understanding and accepting.[189] What is needed now is therefore a characterisation of these circumstances in which epistemic vigilance mechanisms are believed to be triggered.

Epistemic vigilance mechanisms are globally believed to be directed at the assessment of i) the source of the message and ii) the message's content. As far as the source is concerned, addressees will typically try to assess speaker competence (one's possessing genuine information) and benevolence (i.e., the intention of sharing that genuine information) and try to look for evidence of the speaker's reliability and overall trustworthiness; assessment of the source in turn is believed require a metarepresentational ability, it supposes the necessity of being able to make assumptions as to the speaker's intentions. As far as message assessment is concerned, the account of epistemic vigilance basically relies on the relevance-theoretic framework: "the search for a relevant interpretation, which is part and parcel of the comprehension process, automatically involves the making of inference which may turn up inconsistencies or incoherencies relevant to epistemic assessment" (Sperber *et al*. 2010: 376). Chapter III of the present work will specifically be devoted to exploring the technicalities of this last assumption, in an attempt to explore the cognitive implications of the link that can be established between comprehension and acceptance (in other words belief).

A final aspect of the epistemic vigilance framework which is of paramount importance here is its link with reasoning and argumentation. Since addressees who are epistemically vigilant will typically look for incoherence and inconsistencies, speakers aiming at convincing their audiences need to make sure either i) that the addressees do not spot such inconsistencies (and this is what this thesis is largely about) or ii) that there are no such inconsistencies in the message. This second option is the point of entry for the inclusion of argumentative concerns in the account of epistemic vigilance; Mercier and Sperber have very recently developed to that effect what they call the argumentative theory of reasoning (see Mercier submitted, Mercier & Sperber 2009, Sperber & Mercier in press, Mercier & Sperber forth.), and argue that reasoning is a tool that has evolved for epistemic vigilance needs, both for the assessment of communicated material and the production of convincing arguments. Evidence from research on the confirmation bias, they claim, actually supports this view as the confirmation bias obtains only in cases of reasoning (while absent in intuitive inference); moreover, if taken as a feature of reasoning rather than as a flaw, the confirmation bias is a mechanism which contributes to the division of cognitive labour in argumentative exchanges: since conversational participants will tentatively only look for evidence confirming their own standpoint and assess other's

[189] See Sperber *et al*. (2010), section 3, for an informed discussion of this link, where the idea that comprehension requires some kind of stance of trust is motivated.

contributions in light of the latter, argumentative discussions are likely to be much more productive than individual reasoning (and this is something that research on the Wason Selection Task paradigm confirms, as people perform much better in groups than individually, see Moshman & Geil 1998). Such a relationship between reasoning and epistemic vigilance, if taken from the perspective of argumentative exchanges, is thus ideally tailored to characterise one of the ways people are able to screen or filter communicated information.

II.3.2.2. Cognitive and emotional filtering

Clément (2006) already laid the grounds for an account of human gullibility based on a dual mechanism of information filtering. His aim is to explain how, despite the emergence and stabilisation of strategies meant to defeat cheaters and deceivers through evolution, we still end up deceived through communication. This leads him to identify not only protection mechanisms, but also the nature and functioning of our information-processing abilities.

The first filter Clément assumes to be working when processing communicated material is cognitive in nature: it assesses i) the nature of the context on which the information is transmitted to the individual (whether the interaction is e.g. a friendly conversation between siblings, a negotiation between politicians of opposing parties, a commercial advertisement, etc.), which will influence the credit the individual will give to the message; and ii) the content of the message, which will be examined in light of already held beliefs and values. Both scanning operations are very quick and will take into account only superficial properties of the source and the message.[190]

The second representational filter Clément assumes to be functioning when processing communicated information is emotional in nature, and is basically geared towards an assessment of how pleasant both external and internal stimuli are to the organism. The idea behind the usefulness of such a conception is that this type of filtering processing can constrain the input that cognitive processes will subsequently take: the organism will presumably grant pleasant stimuli more attention that unpleasant ones, and this is something that may explain how certain information is left unprocessed; in turn, this constitutes an advantage for an account of manipulative communication, to the extent that the latter specifically aims at leaving certain information out of the set of information the successfully manipulated addressee will process. Such a conception is reminiscent of something Tanaka (1994) postulated, namely that pleasurable stimuli act as 'attention-grabbers', as it were, so that the information they

[190] Note that this is already a first formulation of the detailed epistemic vigilance account found in Sperber *et al*. forth.

accompany is more likely to be processed: as Tanaka notes, in covert communication "the communicator relies on the addressee noticing certain non-linguistic stimuli, given the way in which her general cognitive system is organised. Sex and food are typical stimuli which draw the audience's attention" (Tanaka 1994: 41). In cognitive terms, a manipulative speaker may expect to be successful if she manages to make certain information salient (to the expense of other information), and this can be achieved by relying on 'emotional' means. What Clément says in this respect is quite instructive:

> "The role of emotions is thus not limited to the motivation or the inhibition of mental processes. They indeed seem to operate a preliminary selection on the objects that will tentatively be later cognitively processed more in depth. By continuously 'scanning' the environment, emotional processes thus bear the function of orienting the subject's attention, both when a stimulus does not meet situational expectations and when an obstacle hinders the realisation of a plan of action. […] Accordingly, emotions actively partake in the constitution of beliefs since the latter depend on the nature of the information mobilised in cognitive processes." (Clément 2006: 217, my translation)

From an argumentative viewpoint, in particular in what regards fallacies, it is very easy to find a counterpart in discourse to these processes. All fallacies relying on emotion, such as the *ad baculum* (appeal to a threat) and the *ad misericordiam* (appeal to pity or misery), typically exploit the target's propensity to be subject to emotional reactions. In a manipulative scenario, one could interpret this as an attempt by the speaker to constrain information-processing by trying to prompt emotional responses which will keep the target's attention away from other information, presumably information about the argument's problematic acceptability or about the speaker's deceptive agenda.

II.3.3. RT AS A SUITABLE FRAMEWORK FOR A COGNITIVE ACCOUNT OF MANIPULATIVE COMMUNICATION

II.3.3.1. RT meets cognitive requisites

Several crucial points have to be highlighted from the sections constituting this state of the art on influence, persuasion, and manipulation. Let me list them here:

- Information-processing mechanisms can be studied and modelled according to experimental findings in cognitive, social and evolutionary psychology.
- Information-processing mechanisms are fallible.
- Information-processing can take different paths (within the limits of systematic processing on one bound and peripheral or heuristic processing on the opposite bound).

- Most biases and processes of influence that occur through communication are prone to receive an explanation in terms of information-processing (notably in terms of relevance and salience).
- Information-processing mechanisms have evolved alongside mechanisms meant to make sure that we do not systematically end up being deceived: epistemic vigilance and representational filters are tailored to prevent this from happening.
- Manipulation, according to the definition given in chapter I (see §I.3.2.), is successful when the speaker manages to make the addressee focus on certain information and at the same time fail to focus on other information. I will assume that this, in itself, can be reinterpreted as the exploitation of the inherent fallibility of epistemic vigilance and representational filters.

Having reviewed these cognitive aspects of information-processing, I am now in a position to refine the formulation of the purpose of the present research, which is to explain the success of manipulative communication, the latter being defined as the situation in which epistemic vigilance and both cognitive and emotional filtering fail to produce the outcomes that should lead to the rejection of the communicated information or of the courses of action implied by the information that has been communicated.

The main task I am facing now is the description and explanation of how this can be achieved through communication. To this end, I will make use of the findings reviewed in this section on information-processing, as I believe there are strong grounds to assume that manipulation is precisely about taking advantage of the inherent fallibility of screening mechanisms which we take to characterise information-processing.

What we need is thus an information-processing model that:

> i) accommodates cognitive considerations as well as linguistic ones; more precisely, we need a model which bridges the gap between cognition and language, between beliefs and words.
>
> ii) accommodates fallibility, cognitive biases and heuristic processing, so that it can explain why we sometimes interpret messages without noticing that the speaker is pursuing alternative goals she does now want us to know about.
>
> iii) explains why processing stops at a given point and why such processing was prompted in the first place.

RT (cf. Sperber & Wilson 1995) can arguably make for all these requisites. Inherent to the model is an account of how people acquire new information, get rid of and revise previously held

information on the basis of contextualised communicated material, which makes an explicit link between cognition and language, so that answers the requisite number (i). RT accommodates fallibility, cognitive biases and heuristic processing, in that it postulates that comprehension mechanisms are heuristic by nature: it can therefore account why and how misunderstandings may occur – and by extension why manipulation can occur. So that should take care of requisite number (ii). Third, RT is probably the only pragmatic theory that explicitly postulates a 'firing threshold' where processing will stop.[191] That means that it formulates criteria meant to explain why it is that the individual will consider at a given time having reached a satisfactory outcome of meaning calculation. Moreover, considerations of relevance are able to explain the selection of input assumptions the individual will mobilise as s/he processes the information, and this is a crucial step in explaining why processing is oriented in one direction instead of another, which is in my view decisive to meet requirement number (iii).

Before I proceed to expose the technicalities of a pragmatic account of manipulative communication, let me add a few reflections about specific pragmatic aspects of information-processing, which can also in a variety of cases be responsible for the success of manipulation.

II.3.3.2. RT can explain pragmatic-processing-related manipulation

Following Allott (2007) and Sperber & Wilson (1995), I will consider that the type of information-processing governing communicative exchanges, i.e. comprehension, is a heuristic, as it "can ignore information, where an algorithmic procedure would have to take all information into account" (Allott 2007: 266). It is automatic, intuitive, and may yield errors, i.e. it is fallible. This is crucial for my account of manipulation, because the latter requires that some explanation of why certain information is missed be provided, and this is typically something that RT can do.

In order to show how RT is naturally suited to host such an account, let me analyse a clear example of manipulation where implicature derivation plays a determining role. The example comes from an advert for a lottery product sold in New Zealand,[192] and pictures a group of university students sitting in an exam room. One of the students, having finished a bit earlier than the allotted time, pulls out of a pocket a lottery ticket, which he starts to scratch in order to see if he won. The supervisor of the exam calls the end of the test and instructs all students to put down their pens and hand in their papers to the front, on a pile. Once everyone is gone, the

[191] The argument was formulated by Maillat in Oswald & Maillat (2010).

[192] "Instant Kiwi scratching cards". Advert available on http://www.youtube.com/watch?v=2NfuniN0Sdg. Last accessed on 02.08.2010.

student who was scratching the lottery ticker realises that he is alone, and goes to the front desk to hand in his test. The supervisor says it's too late, and the following exchange takes place:

(66) a. Supervisor: "Sorry, too late. Gave you plenty of warnings about time. You failed. Sorry"

b. Student: "Excuse me, do you know who I am?"

c. Supervisor: "I have absolutely no idea"

d. Student: "Good" (and puts his test in the middle of the pile of papers on the desk)

Here the student manipulates the supervisor by letting the latter derive an implicature which would otherwise be warranted in the context, only that he does not signal in any way that the derivation of the implicature is instrumental to the fulfilment of his manipulative goal, namely to hand in the paper and have it corrected even if he turned it in late. Assuming hierarchical roles in the situation (teacher *vs*. student) to be mutually manifest to both conversational participants, assuming that the student is 'faulty' as he wants to turn in his paper late, it is reasonable to expect him to negotiate the acceptance of his test by the supervisor; moreover, and more crucially, (66b) could perfectly be taken as a way of implying that the student has relations that might get the supervisor into trouble, so it could be taken as a threatening speech act. Prosodic pointers in the ad seem to furthermore confirm this assumption, as the student adopts a somewhat superior and disdainful tone. Thus, in (66c) the supervisor is answering the threat by signalling by his ignorance of the student's name that he is not intimidated. And this is precisely the type of answer the student covertly intended the supervisor to produce. It allows him to confirm that he is safe in putting his test somewhere in the middle of the pile, as the supervisor does not know his name and will thus be unable to sanction his late test delivery.

The manipulative twist of this example lies in the fact that the supervisor has wrongly mobilised contextual assumptions which would be relevant to the threat scenario, but which are irrelevant to the explicit meaning of the student's question in (66b). In other words, the calculation of meaning has led the supervisor to adopt an erroneous belief about the conversational point, namely that the interaction was about a threat, while in fact it was just an information-seeking interaction that the student was covertly hoping to fulfil. Given the information available to both participants at that time and the mutual awareness that a conflict is in progress, the assumption according to which the student will try to work his way out of the conflict to his own advantage is highly accessible, thus easy to process, and therefore relevant, which explains why the supervisor kept it in the meaning derivation calculation. This assumption, combined to (66b), expectedly leads the supervisor to interpret (66b) to be tantamount to a threat. It should also be added that (66b) is typically the type of utterance one finds in literature, movies, or TV series

when a character tries to gain some advantage by putting forth her/his personal connexions; in this sense, also, (66b) could be taken to function as an *ad baculum* argument, even if it is only an apparent '*ad baculum*ness'.

This example shows that comprehension can be abused to instil false beliefs and manipulate people to one's own advantage, and it is a vivid illustration that reliance on contextual factors determined by the situation coupled to particular wordings which are symptomatic of widely known situations (that can be interpreted as culture-specific scenarios), can constrain the interpretation of the message in deceptive ways. It was very unlikely, in the context of the advert, that the literal interpretation would go through as the most relevant one, precisely because all other contextual information prompted for a richer interpretation, and thus, presumably, for a more relevant one (compatible with the threat scenario).

The fact that the success of manipulation here rests on the 'undue' derivation of a target implicature – which is actually irrelevant to the goal the student is pursuing – is evidence not that information-processing, in terms of comprehension, can go wrong (because in another scenario where the student is indeed formulating a threat, what the supervisor understood would be the intended meaning), but evidence that information-processing can be constrained so as to engage the addressee on a track which has no relevance to the manipulator's actual manipulative intention. I will show in Chapter III how this can be achieved.

There are grounds to discuss whether the comprehension heuristic yielded an error in example (66), to the extent that (66) is not a case of overt communication: the student is pursuing a goal he is silent about, so the real point of his utterance is not made mutually manifest and hence cannot enjoy communicative status (and cannot be said to yield any presumption of relevance whatsoever). Accordingly, given the context, the content and the form of what he utters, one would be hard-pressed to consider that interpreting (66b) as a threat is an error, precisely because our encyclopaedic knowledge tells us that what is tentatively made mutually manifest in this sort of conflictual context is the relevance of a threat. Therefore, the supervisor is not wrong in interpreting (66b) as a threat – he would be if the goal the student made mutually manifest was overtly different. Saying that he is wrong would mean that he misunderstood the student's intention, and this in turn would mean that the student made his intention mutually manifest. But this is not the case, and it is precisely for this reason that accounting for manipulation in pragmatic terms is so finicky; in any manipulative exchange there is a level of communicative meaning where information-processing on behalf of the manipulated addressee follows, rightly, the standard procedure and a level of meaning which remains concealed and therefore about which no processing can occur (unless the addressee's epistemic vigilance mechanisms are triggered and identify that something is fishy in the utterance).

Let me add a final example to motivate the choice of RT as a suitable framework to account for manipulative communication. Many puns and mind twisters rely on specific pragmatic properties of information-processing to be successful and in this respect they can be put on a par with manipulative communication taken more broadly, because they all suppose that some information is left unattended by the addressee at some point.

Imagine someone asks you the following question:

(67) "Some months have 30 days and some months have 31 days. How many months have 28 days?" (taken from Fixx 1977: 80)

The expected answer is "One: February". However, the correct answer we are looking for is "all of them", because it is indeed true that all months have 28 days, even if most of them have more. What happens here is that the first two utterances, via processing of scalar implicatures, each call for the contextual construction of a subset of the set of 12 months within which some months will be singled out according to their exact duration, that is, a subset of months counting exactly either 30 or 31 days, which prompts the addressee to understand that he is being asked about the exact duration of a month. The implicatures one derives from the first two utterances constrain further processing by inviting the hearer to expect that the answer to the question is consistent with the type of interpretation yielded so far, so it would be somehow as if scalar implicature processing is primed through the first two utterances, and this is not surprising since the only interpretation consistent with "some months have 30 days and some months have 31 days" is the interpretation according to which exact month duration, in terms of days, is what is being assessed. So here the constraint lies in that the hearer is led to expect no variation in the type of interpretative procedure to be followed.

What happens in (67) is thus that one interpretative strategy is enforced by being made salient in the context (scalar implicature derivation); as a consequence, this drives the addressee's attention away from the interpretation of the question that would be relevant to what is actually being asked. Instead of interpreting "at least 28 days", the hearer is led to infer "exactly 28 days" simply because this was the point of the two preceding utterances.

Through the account of implicature provided by RT we can easily explain this phenomenon in terms of constraint on interpretation, as I have briefly tried to expose. Chapter III will develop this idea at length through the notion of Contextual Selection Constraint (CSC, cf. Maillat & Oswald 2009, forth.). What we could highlight in the meantime is that this type of phenomenon relates to what is sometimes referred to as a pragmatic illusion (cf. the Moses illusion, see §II.3.1.3.1. above); Allott notes in this respect that the problem with them "is not that the pragmatic faculty does not have potential access to the right information but that it reaches a

decision without considering some highly relevant information that it could have considered" (Allott 2007: 272). In this sense, we can say that the success of manipulative communication crucially rests on its ability to trigger shallow-processing of the information contained in the message.

With these last two examples I conclude this state of the art. In the following chapter I will try to show how a cognitive model of communicated information-processing can articulate the findings of all these different disciplines in order to account for manipulative communication.

CHAPTER III. CONTEXTUAL SELECTION CONSTRAINT: TOWARDS A COGNITIVE PRAGMATIC ACCOUNT OF MANIPULATIVE, FALLACIOUS AND UNCOOPERATIVE COMMUNICATION

"Mais le contexte est plus fort que le concept". Claude M'Barali (MC Solaar)

III.0. INTRODUCTION: MANIPULATION AND COGNITIVE ASPECTS OF INFORMATION-PROCESSING

I have tried in the previous chapters to motivate a cognitive pragmatic account of manipulation and to specify the requirements such an account should meet. For that, I have first defined what I take manipulation to be, before proceeding to a review of literature from different research domains that arguably provide significant contributions to its study. The principal idea that emerged from these works is that of *constraint*: manipulation can be conceived of as operating a constraint on the cognitive processes that are taking place in its target's mind, whereby the success of manipulation rests of the target's derivation of a preferred – manipulation-compatible – output of information-processing. The gist of the account I will now turn to develop lies precisely in this idea, namely that manipulation is a constraint on information-processing, but a constraint of a certain kind which calls for thorough specification, since, quite trivially, expecting someone to interpret any message can be viewed as manipulation to the extent that the speaker is always intending a particular message to come across and modify the hearer's cognitive environment. Since the latter does obviously not match the requirements of the definition I have provided in chapter I (see §I.3.2.), I will need to qualify the type of constraint I take manipulation to enforce, and in order to do so will defend an RT-based technical account to cater for these requirements.

I have ventured that manipulation is a twofold process.[193] The manipulator aims at prompting for a specific interpretation of the message and at the same time aims at getting the addressee not to take into account other (relevant) information, that is, simply fail to process such information, or, at least, fail to find it relevant. In terms of information-processing and following Sperber and Wilson's fundamental assumptions on cognition, this means that the preferred interpretation should appear as relevant for the hearer and that the set of assumptions the manipulator expects the hearer not to process should either appear completely irrelevant in the exchange or be left unprocessed altogether. These conditions can be expressed in the form of two technical requirements any manipulative attempt in communication should meet: i) cognitive processing of whatever set of assumptions the manipulator aims to convey must require little processing effort and yield significant cognitive effects; and ii) whatever set of assumptions that is detrimental to the success of the manipulative attempt must either require too much processing effort and yield no significant cognitive effects or not be represented at all.

This leads to the formulation of a hypothesis regarding operational aspects of manipulation: the success of manipulation is a function of a constraint both on relevance and on irrelevance. The

[193] See also Maillat & Oswald (2009, forth.) for a development of this idea.

message itself, which I call 'manipulation-compatible' because it is the content that the manipulator intends her hearer to derive, should be made relevant while other (potentially equally relevant) information should be made irrelevant or concealed so that its chances of being taken as input for further cognitive processing are lowered to a minimum.[194] As put by Allott in his discussion of pragmatic illusions, those are cases where "the pragmatic faculty reaches a decision without considering some highly relevant information that it could have considered" (Allott 2007: 272). The remainder of this chapter will thus try to discuss how the success of manipulation will depend on whether the manipulator is able to render the manipulative message highly relevant and the problematic information irrelevant.

Perhaps the most interesting aspect of this hypothesis lies in its explanatory potential, because one can actually start explaining in terms of processing effort and cognitive effect how a message can be made relevant or how someone may attempt to prompt for the irrelevance of certain information in a given context through information-processing constraints. This type of phenomenon has been described at length as it has reportedly been shown in cognitive psychological research on biases and illusions; but with an RT-based account, we can start to explain why these biases and illusions work in communicative scenarios.

There are in principle two ways of constraining information-processing: either you constrain the input so that only certain assumptions are processed or the type of processing itself (though both could be constrained at the same time). Constraining the input is about making sure that only a certain set of assumptions ends up selected as input of the meaning derivation procedure (thus leaving other assumptions out of the picture); constraining the type of processing itself is about trying to limit cognitive processing, for instance to a heuristic-based treatment so that it remains fast, automatic and so that the addressee is tentatively less prone to further process the stimulus.[195]

---

[194] This of course is any information that would be detrimental to the manipulative attempt: assumptions about the speaker's intentions, internal inconsistencies, etc., in sum, any type of information epistemic vigilance should identify as problematic in the communicative exchange.

[195] Let me make an analogy: if I want to bake a cake, I will need several ingredients; whether I use wheat flour or peanut flour is going to produce changes in taste in the final product; in other words, the input I use in my preparation will determine the output, and changes in the input will result in changes in the output. Now, imagine that instead of mixing the ingredients for 2 minutes I mix them for 4 minutes, and that instead of baking it 20 minutes in the oven, I put it in for 45 minutes: the output will also be different in both cases, but this time it has to do with the way I prepare the dish, i.e., with 'processing'. I take it similarly that the input to cognitive processes and the type of processing involved yield changes in the output.

If we develop a little further the argument about constraining the type of cognitive processing, in dual-process theory parlance, this amounts to saying that a manipulator is tentatively better off trying to exploit the peripheral than the central route; this would be a simple consequence of the nature of both routes and at the same time of the manipulative requirements mentioned above: the more resources the hearer will allocate to the message, the more issue-relevant thinking he will indulge into, and the more he will be likely to process information which is detrimental to the success of the manipulative attempt.

Let me here add a few more remarks to the discussion about what dual-processes theories may offer to the study of manipulative communication (see §II.2.2.1.), as I believe there is room to discuss whether manipulation should be restricted to peripheral processing or if it could also be successful should processing go central. I have already mentioned that manipulation will tentatively be more effective if peripheral processing instead of central processing is triggered, the reason for this being that its automaticity and 'heuristicity' should in principle decrease the likeliness of the addressee's processing more information in further detail – in particular information about the speaker's motivation or tentative inconsistencies that would become noticeable should the expenditure of processing effort increase. While I do not question the intuitive soundness of this assumption, I think that one could still argue that manipulation can be successful even in cases where central processing is involved, and I can think if three reasons to support this claim.

First, central and peripheral processing are not mutually exclusive processes by definition (O'Keefe 2002), so in principle there is no *a priori* reason to associate the success of manipulation to exclusive reliance of peripheral processing, even if the configuration is somewhat different in manipulative and persuasive scenarios due to the covertness feature required by the first but not by the second. Nonetheless, I see no reason why this should change anything, to the extent that a manipulative message can also manifest apparent overt persuasiveness. It is mainly because manipulation takes the appearance of standard communication, even if it exploits a covert level of intention in addition to that, that it is still analysable with the tools theories of persuasion have to offer us.

Second, there is anecdotal evidence for successful deception even when central processing is involved: in order to show this, let me take the example of a specific riddle that I will call "the missing franc story".[196] The narrative goes like this: three friends go to a restaurant and each of

[196] A very similar version of this riddle called "The missing dollar" appears in (Fixx 1977: 71). In only recently came across it and had heard the 'franc' version years ago, which is why I did not refer to Fixx in the first place.

them orders a soup. They eat their respective soups, but at the end of the meal, when presented with the bill, which is 15 francs (5 francs for each soup), the customers complain about the meal being too expensive, given that one can buy instant soup such as the one that was served for less than a franc in a supermarket. The waiter reports the information to the owner, who decides to cash in 10 francs and returns 5 francs to the waiter telling him to give back their money to the customers. Thinking he can make profit out of that situation, the waiter keeps 2 francs in his pocket and refunds each customer 1 franc. All in all, therefore, the customers have paid 5 minus 1 franc, i.e. 4 francs each. Now comes the tricky question:

(68) The three customers have each paid 4 francs (5 minus 1), i.e., a total of 12 francs (3 times 4), and the waiter kept 2 francs, so that's a total of 2 plus 12 francs, which equals 14 francs: where is the missing franc?

People who are told this riddle usually end up quite puzzled, because indeed there is a 1 franc difference between the initial amount the customers paid (15 francs) and the result of the calculation that leads to 14 francs (12 francs paid plus the 2 francs in the waiter's pocket). And this is what prompts subjects to engage in central processing: it is motivating *per se* to try to find the solution to this problem, and for this cognitive resources need to be allocated to calculation and issue-relevant thinking. As long as they do not come up with the solution, it can be said that manipulation has worked since it means that they have fallen victim for the trick. This is so for a crucial reason: they are actually trying to overcome an inconsistency that they came to know – and furthermore straightforwardly accept – as the output of their own comprehension procedure (which was in turn prompted for by a bias introduced in the narrative).

Before I suggest a pragmatic explanation of this riddle, let me first provide the solution. The question in (68) is misleading in that there is no missing franc in the first place: each customer has paid 4 francs, which is accurate data; the waiter has kept 2 francs in his pocket, which is also accurate data. Now, the owner has cashed in 10 francs (since he returned 5 francs out of the 15 that he initially received). So if we consider the amount of money that stays in the restaurant when the customers leave, it appears that it exactly amounts to 12 francs (10 for the owner, 2 for the waiter), which is obviously exactly what the customers have paid (3 times 5 minus 1 franc, i.e. 3 times 4 francs, i.e., 12 francs). So the trick lies not in an error of calculation, but in the way the problem is presented and framed: after the owner returns 5 francs, there is no further relevance in considering a total of 15 francs, so this number should be eliminated of the equation altogether. Yet, the question "where is the missing franc?" – which presupposes that there actually is a missing franc – taken together with the way the calculation is presented actually maintains the relevance of the information that 15 francs have been cashed in: the speaker wouldn't even speak of a missing franc if no prior difference had been noticed (and thus

presented as accepted by the speaker). In other words, people have trouble with this riddle not because it involves complicated calculations, but because the way the question is phrased misleads them into believing that two pieces of information are relevant to the explanation: i) the information that "an amount of 15 francs has been cashed in" is still relevant, and ii) the information that "a total of 14 francs" is made relevant by an explicit calculation, thereby prompting subjects to fail to notice that the calculation "12 plus 2, that's 14 francs" should actually be ruled out because it adds two numbers which should not be added: the 2 francs the waiter puts in his pocket are already part of the 12 francs paid by the customers, the other part being the 10 francs cashed in by the owner.

What the example shows is that even when people engage in central processing they still might be deceived, and I contend that this is because they fail to question their interpretation of the narrative and the question. Those who succeed in solving the riddle are precisely those who are able to question what they have understood and the accuracy of the calculations presented in the narrative. In other words, central processing can manage to counter a manipulative attempt when it is able to override the results of the comprehension procedure. And since the comprehension procedure can be described in terms of a heuristic (following Sperber & Wilson 1995 and Allott 2007), I suggest that manipulation can succeed under central processing conditions as long as the latter does not override peripheral processing of heuristically-derived input (i.e. as long as it does not call into question the hearer's interpretation of the message).

This last idea is compatible with latest work on reflective and intuitive inference (see Mercier & Sperber 2009) and constitutes a third argument in favour of a central-processing-compatible account of manipulation; reflective inference – or reasoning in Mercier & Sperber's terms –, which characterises reasoning (as opposed to intuitive inference), requires input just as any other cognitive operation in order to unfold. One of its crucial features is that the input it takes is typically the output of the comprehension procedure, which is itself of a heuristic nature. Taking this into account, there is a strong case to be built against the reliability of central processing, not because central processing is itself fallible *by nature*,[197] but because central processing takes as input propositions whose reliability is already questionable as they result from heuristic processing, which is itself already likely to yield inaccurate outputs.

So to come back to the question of knowing whether the characterisation of manipulation should be limited to the exploitation of heuristics, I believe the above discussion allows me to respond negatively: it is quite clearly established that both central and peripheral processing are

[197] This is an altogether different issue.

compatible with the success of manipulation. However, I am still in no position to claim that manipulation can be successful by exclusive reliance on either peripheral or central processing.

Independently of the implications of the preceding discussion for the issue of central *vs*. peripheral processing, the example provided above emphasises once again the importance of the role of comprehension in manipulative attempts,[198] and since comprehension is heuristic in nature, there is a strong case to be built in favour of a systematic study of how manipulation may exploit these heuristics. I have already mentioned a few examples in which the comprehension of communicated material might constrain the addressee's response and influence the success of the manipulative attempt (the Moses illusion, different kinds of biases and fallacies, etc., see Chapter II.3.). In this chapter I will detail my analysis of communicative biases and heuristics that may be abused to further manipulative goals, so I now turn to explore more in depth the technicalities of my theoretical assumptions by first exposing the main challenges faced by a pragmatic theory and then proposing an information-processing model of manipulative communication couched in a Relevance-theoretic framework.

III.1. MAIN CHALLENGES FOR A PRAGMATIC ACCOUNT OF MANIPULATION AND NON-COOPERATION

Now that I have examined and defined in detail my object of study and its relationships with neighbouring phenomena (see Chapter I) and that most relevant literature has been assessed in an attempt to identify the different types of constraints that could make for the success of manipulation (see Chapter II), I can discuss the suitability of a pragmatic framework for an account of manipulative communication.

I have noted that manipulation i) occurs in and through communication, and ii) that its success is to a large extent determined by the way the targets of manipulative attempts process the information contained in the manipulative utterances the speaker uses to fulfil her covert perlocutionary goals. I therefore assume that in principle, there is room to consider that pragmatic accounts of communication, which examine how and why people understand each other, can have much to say about communicative phenomena such as manipulation. However, we will shortly see that major reservations could be formulated against such an endeavour. It is the purpose of this section to voice these concerns and formulate the problem so as to evaluate whether there still are grounds to privilege a pragmatic approach.

---

[198] Recall that this is something that Sperber *et al*. (1995) have already postulated in their pragmatic account of the Wason Selection Task (see §II.3.1.1.1.).

III.1.1. MANIPULATION AND SPEECH ACT THEORY

If we want to consider manipulation as a communicative speech act, we could envisage that its treatment should be akin to the treatment of other speech acts within the scope of SAT. But such a standpoint seems unsatisfying for the following reasons.

Firstly, if manipulation is indeed a speech act, it falls outside the scope of the principle of expressibility.[199] I have earlier characterised manipulation (see §I.3.2.1 and §I.3.2.2) as something the success of which necessitates that it remain concealed. In the case of manipulation, thus, the speaker must not use any linguistic form which actually expresses what she actually intends to achieve (i.e. that she is trying to manipulate her hearer). Hence the meaning of the speech act of manipulating cannot be expressed the same way the meaning of, say, a promise, can. To complement this objection, Parret (1978) notes that a manipulative speech act cannot be prefaced with an explicit performative verb like any other speech act, for the same reasons: one cannot utter something like "I manipulate you (by saying) that P", while one can always say something like "I promise you that… P".[200] And thus to consider manipulation as a performative speech act is self-defeating (cf. Parret 1978: 15-16).

A number of observations should be made on this particular issue. First of all, it should come as no surprise that manipulation escapes the principle of expressibility, for the very nature of manipulation supposes that it should remain undisclosed. Manipulation does not bear a communicative status, that is, no manipulative speaker intends her hearers to recognise that she is trying to manipulate them. A direct consequence of this is that manipulation is not an illocutionary act *per se*, since it is inavowable (Strawson 1964: 454). If we can assume, within a Speech Act theoretic account, that unless the hearer grasps the illocutionary force of a speech act, the speech act is not successful, we cannot say the same of manipulation, because no expectation of illocutionary uptake comes with it.

From these observations emerges the idea that the act of manipulating is a *perlocutionary* act, not an *illocutionary* one. As such, and just like persuading, surprising, dissuading, convincing, etc., manipulation denotes a consequence – or an effect – of a given utterance on the hearer's

---

[199] The principle of expressibility expresses the idea that we can always say what we mean, i.e., that there always is a linguistic form that can be uttered and which corresponds exactly to what the speaker means: "We might express this principle by saying that for any meaning X and any speaker S whenever S means (intends to convey, wishes to communicate in an utterance, etc.) X then it is possible that there is some expression E such that E is an exact expression of or formulation of X" (Searle 1969: 20). This principle turns out to support the claim according to which the rules governing the *performance* of speech acts are equivalent to the rules governing the *utterance* of certain linguistic elements (Searle 1969: 20-21). In other words, the principle of expressibility entails the idea that uttering a speech act is performing it.

[200] A similar observation is made in Blass (2005: 170).

beliefs and/or behaviour, and not what the speaker manifests her utterance is really about. This means that we cannot, at least with the tools SAT provides, directly link manipulation to speaker meaning, as according to the elements discussed in what precedes there is no properly manipulative meaning, 'meaning' being understood here as a set of assumptions the speaker intends to communicate to her audience. There is consequently no point in trying to analyse manipulation in terms of the rules governing the production and reception of standard speech acts, namely their felicity conditions.

Felicity conditions denote the conditions that must be met in order for an illocutionary speech act to be successful, among which we find the sincerity condition (requiring the speaker to be sincere with respect to any implications her utterance may carry) and the essential condition (which is related to the conceptual representation of the speech act). Searle's classification of the latter is helpful when it comes to characterising certain types of manipulation; for instance he gives the example of insincere promises, which are taken to play on the level of the sincerity condition, phrased as follows: "S [the speaker] intends that the utterance of T [the sentence] will make him responsible for intending to do A" (Searle 1969: 62). The particularity of an insincere promise lies in the fact that the speaker manifests her commitment to fulfilling whatever she promises, while in fact she is secretly not committed to it. It could be tempting to jump to the conclusion that manipulative utterances systematically violate some felicity condition of the speech act, but this solution is inadequate given i) the impossibility of considering manipulation as an illocutionary act, and ii) the fact that some cases of manipulation (such as the Jo Moore scandal, see §I.2.3.2.1) do not break any felicity condition, since in such cases the manipulation is not about misleading the hearer, but about counting on him not to notice information which is otherwise perfectly unproblematic.

Incidentally, we face once again the problem I noted above about the difference in nature between the manipulative intention and other types of intention that partake in communicative phenomena (see §I.3.1.5): while the manipulative intention is successful only when it is not recognised, other intentions pertaining to the speaker's meaning are successful and realised when they are indeed recognised by the hearer. Taillard (2000) summarises this point in a very clear way, as she takes intentions of the sort of the manipulative intention to concern the perlocutionary level of communication:

> "All that is required for successful communication is the performance and identification of an illocutionary act – whereby the hearer will understand a speaker's utterance, the normal states of affairs will be brought about, and appropriate responses will be 'invited'. Perlocutionary acts are peripheral to the study of communication proper." (Taillard 2000: 147)

Thus, the question of assessing whether we can use SAT to account for manipulative communication runs the risk of posing a false problem, to the extent that we cannot indeed treat manipulative speech act as any other standard speech act; I hypothesise in order to support this claim that there does not seem to be any manipulative illocutionary force, or that if there is one, then it is radically different from all other illocutionary forces which have a communicative status, i.e., which are avowable. It is thus not that great a surprise that SAT's theoretical apparatus shows difficult to apply to a phenomenon such as manipulation, simply because it seems to fall outside of the scope of phenomena SAT was designed to account for.

III.1.2. MANIPULATION AND GRICEAN FRAMEWORKS

Through my discussion of IMT (see §I.2.3.), I have come to a number of conclusions regarding the adequacy of any Gricean framework for an account of manipulation. Let me elaborate on these.

I have shown that Gricean theory does not provide strict criteria for discriminating between intentional and unintentional violations of the maxims (in the case of problematic utterances such as manipulative ones). Grice's theory of communication builds on the assumption that interlocutors understand each other on the basis of their mutual recognition that certain standards (i.e. conversational maxims and the CP) are being met – and this includes their overt exploitation –, since they are supposed to act cooperatively. With such a specification, all conversational moves which *covertly* fiddle with these standards can arguably be taken to lie outside the scope of the model. So just like SAT appears not to be able to deal with manipulation by virtue of the latter's non-requirement of hearer uptake in order to be successful, Grice's model similarly seems unequipped to deal with phenomena which it was not meant to capture.

Manipulation by definition is a communicative situation where standards of communication do not hold, in particular from the factual perspective of the manifestness of the goals interlocutors are actually pursuing, which makes its hosting within the Gricean model problematic.

This concern relates to the concerns I voiced earlier about the intentional nature of manipulation (see §I.3.1.5.). Even if a manipulative utterance is intentional, it cannot be said that it is intentional *in the same exact way* as any regular, cooperative, utterance is intentional. There is a crucial difference to be acknowledged between an intention, which can remain private notwithstanding its covert instantiation in a conversational exchange, and an intention which is meant to be identified and *recognised as being meant by the speaker to be identified by the*

hearer.[201] While according to contemporary pragmatic models all communicative contents traditionally display what we could call 'manifest intentionality', such is not the case of manipulative intentions. These are meant to be realised, but their fulfilment does not rely in any way on their being recognised as intended.

This is the major challenge any pragmatic theory of communication construing conversational exchanges through the lens of intention recognition will face. The question could be presumably devastating for such approaches: since manipulation requires that the intention it is embedded in remain concealed, what kind of contribution could a theory positing intention recognition as a crucial element of explanation possibly offer?

Perhaps a way to accommodate a pragmatic perspective with the problematic nature of the deception would be to start off from this relatively trivial observation: the success of manipulation still rests on the success of the communicative exchange of meaning it operates on. When manipulation takes the form of a conversational exchange, it is still a genuine instance of communication, in so far as meaning goes through. Let me take the milk example again (reproduced and slightly adapted in (69) below) in order to develop this idea; imagine I want my flatmate Laszlo to leave the flat we share for a half hour because I need to do something I do not want him to know about (say, set up a birthday surprise for him), and that in order to drive him away I utter (69):

(69) Could you go to the grocer's and get a few cartons of milk? We've ran out of milk and it would be nice to have some for tomorrow's breakfast.

(69) makes perfect sense, and in case I actually utter it without the manipulative intention of sending Laszlo away, it could constitute a perfectly cooperative contribution to an exchange. We might say, following IMT, that what would make (69) deceptive is a violation of the maxim of quality. But this is very questionable, since it might still be argued that I'm not uttering any false information in the case we are actually out of milk and that I indeed would appreciate having some for breakfast the next morning (even if my motivation for uttering (69) happens to be that I want to be alone in the flat). The idea is that I could entertain those thoughts and desires and still, independently, want to manipulate Laszlo: that scenario, nonetheless, remains compatible with a construal of (69) as manipulative, because the absence of milk and my desires to have some for breakfast the next morning are effective states of affairs. At the level of what is said, everything looks moreover normal: I am formulating a request and providing a relevant argument in support of that request. At the level of what is implicated, there is not much to say,

[201] This obviously refers to, albeit in a different formulation, to Sperber & Wilson's notion of *communicative intention* (Sperber & Wilson 1995).

since (69) does not appear to convey obvious implicatures, except for contents related to what Laszlo might infer my motivations for not going to get milk are (laziness, being too busy, etc.). In both scenarios, thus, it appears that a violation of Quality is not necessary for the manipulative attempt to go through. Again, the problem in accounting for (69) relates to the fact that Grice's model is meant to tackle speaker meaning as manifested ostensively, not higher-order intentions which rely on the recognition of speaker meaning by the hearer.

Another account of (69), in the Gricean vein, would be to posit a violation of Relation, since the argument I'm putting forward is not directly relevant to my intention of getting Laszlo out of the flat: I'm sending him to buy milk primarily *because I want the flat for myself*; even if it is true that we are short on milk, I am not sending him out *because* I need milk. It is in this sense that we could potentially interpret (69) as involving a violation of Relation. Yet, the violation of the conversational maxim is relative to my goal of getting Laszlo out of the flat without him knowing, and this can certainly not constitute a *shared direction* in the exchange; as my contribution is in addition truthful, neither can it be said that it constitutes an obstacle to the shared goal I am making manifest, namely to secure that we have milk for breakfast next morning (which is the goal I am indeed making manifest with my request). The crucial point here, which emerges as we try to explain this particular instance of manipulation in terms of maxim violation, resides in the fact that manipulation seems to involve two separate intentions (or levels of intentionality), each related to a separate goal: a first communicative[202] intention geared towards the satisfaction of the goal explicitly instantiated (getting milk for tomorrow's breakfast) is instrumental to the success of a second intention which is not communicative *per se*, but which is determined by the hidden motivation of driving Laszlo away from the flat. I believe those are two very different intentions relating to two different goals, and I would venture that in the example manipulation can work precisely because Laszlo's getting the milk is compatible with both.

The apparently paradoxical situation could be characterised as follows: depending on how we look at it (i.e. depending on the goal we choose to take into account), (69) is at the same time cooperative and uncooperative. The utterance, if taken to serve the perlocutionary goal of fulfilling an adequately motivated request, is unproblematic and observes communicative standards. If manipulation is moreover successful, that is exactly the nature of the exchange Laszlo will believe he took part in. Yet, the moment we take the exchange to serve an alternative perlocutionary goal, that of getting Laszlo to leave without letting him know that such is the real

[202] Here I use 'communicative intention' to denote an intention involved in the successful overt exchange of meaning, i.e. one that is about communicating something, and not in Sperber & Wilson's specific sense.

motivation of the request, manipulation arises. In other words, it seems that manipulation lies not in the message itself, but in the intention with which the message is produced, which incidentally is something IMT cannot capture, as it focuses primarily on properties of messages. More importantly, whether there is a manipulative intention or not is not a linguistic problem at all, since the issue is not so much the actual meaning of the utterance, but the function the utterance serves, and more importantly, the function it serves *with respect to a certain goal*.

Now that we have established that manipulation involves two simultaneous levels of intention (on the one hand a completely straightforward level of overt communication, and on the second hand a level of covert intentionality which takes advantage of the first), we need to make explicit where this leaves us in terms of the type of model we should use to account for it.

We have seen that a pragmatic model *à la* Grice, which places the recognition of overtly communicated intentions at the centre of communication is in principle underequipped to deal with manipulation, since the success of the latter precisely requires that the manipulative intention remain concealed (and thus not recognised). This is something I have tried to show by criticising the explanatory adequacy of IMT. The main problem I identified basically boiled down to the fact that any Gricean-inspired model is inadequate to deal with manipulation and related phenomena where concealment seems to play a decisive role because those phenomena lie outside of the scope of the model. They are plainly uncooperative, in that at least one of the goals of the exchange *is not shared* among conversational participants. Any model stating that conversational participants can be described as adopting a behaviour which is consistent with a mutually accepted direction in the exchange will fail to capture manipulation *because there is no mutually accepted direction in the exchange* – at least no mutually accepted direction directly linked to the speaker's goal of misleading the hearer into adopting a problematic belief or behaviour.

So it all comes down once again to the very definition of manipulation, since its success depends on its ability to conceal its manipulative intention. The crucial difference with cooperative communication could consequently be identified as a difference in intention. While cooperative overt intentions prompt hearers to engage in cognitive processing intended in part to figure out the intention motivating the production of the utterance, covert intentions do not. On a technical note, while one can derive explicatures and implicatures to reach a representation of the intentions the speakers tries to make manifest, one cannot derive explicatures and implicatures in the same way to figure out a speaker's concealed intention, since there is no evidence whatsoever that these have been intended by the speaker to be derived in the first place.

Now, such an analysis could be taken to mean that we need to abandon Pragmatics altogether as a possible theoretical framework for accounting for manipulation. However, there is a way of

salvaging pragmatics, and I now proceed to explaining why as I expose a few aspects of Relevance Theory which may lead to the conclusion that valuable and informative research on manipulation is indeed conceivable within the pragmatic domain.

III.1.3. MANIPULATION AND RELEVANCE THEORY

III.1.3.1. Ostensive-inferential communication

Relevance Theory (Sperber & Wilson 1995) is a naturalistic and mechanistic pragmatic theory of communication designed to account for the way people understand each other. It is naturalistic to the extent that it explicitly sets out to account for our natural communicative abilities and in this sense aims at providing a cognitively and psychologically plausible model of human communication. It can be characterised as mechanistic to the extent that it conceptualises phenomena of comprehension as inferential procedures which can be identified and described according to specific cognitive principles and communicative standards governing the way we process information when its source is a message produced by another human being. It applies to verbal and linguistic phenomena but also hosts a more general model of communication that captures non-linguistic communication as well, as it builds on the Gricean idea that a way of conveying information "is to provide direct evidence of the one's intention to convey it" (Sperber & Wilson 1995: 23). Hence, this model assumes that more crucial to the possibility of communication is our ability to make manifest and recognise intentions than the communicators' mastery and usage of a common code. What is decisive for communication thus is not so much the code itself, but the fact that

> "hearers are interested in the meaning of the sentence uttered only insofar as it provides evidence about what the speaker means. Communication is successful not when hearers recognise the linguistic meaning of the utterance, but when they infer the speaker's 'meaning' from it." (Sperber & Wilson 1995: 23)

If follows that Sperber & Wilson's model strongly emphasises the crucial role of *inference* in the success of communication; yet, it does not refrain from stressing that inference, in case it is elicited from and within verbal communication, must rely on coding-decoding processes, for individuals who are communicating by verbal means need to master a common linguistic code in order to convey thoughts through linguistic stimuli.

Human cognition, according to Sperber & Wilson, is geared towards the task of achieving an accurate representation of the world:

> "It seems that human cognition is aimed at improving the individual's knowledge of the world. This means adding more information, information that is more accurate, more

> easily retrievable, and more developed in areas of greater concern to the individual. […] Thus, long-term cognitive efficiency consists in improving one's knowledge of the world as much as possible given the available resources." (Sperber &Wilson 1995: 47)

Clément (2006: 15) notes schematically that there are three sources from which we can acquire beliefs: perception, natural inference and communication. In the case of communication, RT hypothesises that since we aim at improving our knowledge of the world, in principle we should be equipped with some kind of natural tendency to pay attention to and ultimately process what people want to say to us. If we go by these assumptions, it follows that there is a clear motivation in *a priori* expecting whatever people may want to tell us to be relevant, as it may contribute to fulfil our quest for knowledge. If we deem a stimulus relevant, that is, if we recognise a stimulus as being ostensibly directed at us, chances are that we will process it; the presumption that we will find relevance in messages which are directed at us hence constitutes one of the crucial underlying assumptions of RT's proposal. This is what the *Presumption of optimal relevance* captures (in part):

(70) "The ostensive stimulus is relevant enough for it to be worth the addressee's effort to process it" (Sperber & Wilson 1995: 270).

From the speaker's perspective, in turn, this means that she can in principle count on her addressee's disposition or willingness to process whatever linguistic stimulus she makes manifest to him, that is, any ostensive linguistic stimulus she produces.[203] Fundamental to RT is therefore the evolutionary assumption that communication has evolved in order to secure advantages (e.g. obtaining relevant information) from which communicators can benefit. Accordingly, Sperber & Wilson formulate a very general principle which applies to every ostensive-inferential stimulus, the *Principle of relevance*:

(71) "Every act of ostenstive communication communicates a presumption of its own optimal relevance." (Sperber & Wilson 1995: 270)

Sperber & Wilson elaborate on the implications of such a construal of communication, as far as the level of intentionality is concerned, by postulating that every ostensive-inferential stimulus comes with two intentions: the informative and the communicative intention. The informative intention, which is meant to cover what is directly related to the meaning the speaker wishes to convey, consists in making "manifest or more manifest to the audience a set of assumptions I" (Sperber & Wilson 1995: 58). The communicative intention is defined as the speaker's intention

[203] Notice here that my earlier characterisation of Communicative Cooperation (see §I.1.2.2) echoes such a formulation.

of making manifest that she wishes to have her informative intention recognised. Communication is deemed to succeed when the speaker manages to get the hearer to be aware of a set of assumptions; one efficient way of securing this, according to Sperber & Wilson, is to make it manifest that the speaker intends her hearer to be aware of that particular set of assumptions. In this, RT echoes the Gricean assumption that one of the ways the speaker can convey certain information is by trying to secure that the hearer recognises that she intends to convey that information.

III.1.3.2. RT's explanatory scope: the cognitive dimension of information-processing

RT supplements its descriptive account of ostensive-inferential communication by furthermore attempting to explain how it works, notably in view of shedding light on the mechanisms by which hearers manage to adequately derive speaker meaning. Unlike Grice, who conceives communicative exchanges as cooperative endeavours where conversational maxims are deemed to operate, Sperber & Wilson develop an account centred on the idea that communication is best explained by cognitive principles.

As they defend a (partly) inferential conception of communication, they need to account for the cognitive dimension of communicative phenomena, and do so by focusing on information-processing and setting out to specify the conditions under which such processing unfolds. Assuming that human beings are "*efficient* information-processing devices" (Sperber & Wilson 1995: 46, my italics), it should seem quite reasonable to also assume that their behaviour (including communicative behaviour) is regulated by economic principles aiming at securing that in order to reach a goal, the organism allocates enough resources (and not more than required). The interpretation of speaker meaning can thus be conceptualised as a goal the hearer is naturally willing to attain by virtue of the presumption of relevance, the attainment of which requires him to allocate cognitive resources to the processing of the speaker's stimulus.

Processing an utterance in order to derive speaker meaning is thus a cognitive operation which requires the expenditure of cognitive efforts spent in order to yield interesting contextual (and cognitive) effects, among which probably the most basic one is understanding what the speaker meant. Contextual effects, defined as modifications of the cognitive environment (i.e. the set of assumptions available to him at the time he processes the utterance) can be of three types: either the output representation adds new information to the cognitive environment, confirms old information, or prompts for the deletion of old information which turns out to be inadequate,

inaccurate or false.[204] The originality of RT resides in postulating that this ratio between effort and effect is precisely what defines relevance. As a cognitive notion, relevance thus denotes a mechanism which can explain and predict processes of interpretation. It follows that relevant interpretations (i.e., those which are likely to be yielded by the interpretative process) are those which achieve the best balance between effort and effect; furthermore the less effort a representation requires to be derived, the more it is likely to be relevant, and therefore the more it is likely to match the speaker's intentional meaning. In parallel, the more contextual effects a representation produces with regard to the hearer's cognitive environment, the more it is likely to be relevant, and therefore, the more it will be likely to match the speaker's intentional meaning. Both observations converge in the idea that the representation with the best ratio between cognitive effort and contextual effect is the one which can be taken to correspond best to the speaker's original intention.

On a more technical note, RT postulates that three cognitive operations characterise the processing of verbal ostensive stimuli: the derivation of the *logical form* of the stimulus, which is a syntactically and semantically structured sequence of concepts that supplies a sort of template for further processing; the derivation of the *propositional form* and other *explicatures*, mainly through the disambiguation and the enrichment of the logical form; the derivation of *implicatures* (both explicatures and implicatures are to a large extent derived from the contextualisation of the stimulus within an appropriate and relevant set of contextual assumptions). The derivation of implicatures, in RT, does therefore not depend on the existence of conversational maxims; instead, RT keeps the idea of relevance (from Grice's maxim of relation) under a cognitive form but gets rid of the remaining three maxims by assuming that they are reducible to that of relevance (cf. Wilson & Sperber 1981), to the extent that the quality, the quantity and the manner with which information is conveyed are properties which pertain anyway to the assessment of the stimulus' relevance in the cognitive environment.

III.1.3.3. Manipulation and RT

If we want now to assess, just like we did for SAT and the Gricean model, whether RT can host an account of manipulative communication, chances are we might face the same type of objection we encountered earlier. RT accounts of ostensive-inferential communication; yet, there is nothing ostensive about the manipulative attempt *per se*. The covert nature of the manipulative intention makes it fall outside of the scope of the type of phenomena RT captures as *ostensive*.

[204] See Sperber & Wilson (1995: 108-117) for a detailed exposition of the notion of contextual effect.

As for the manipulative intention, let me recall that it is different from the informative and the communicative intentions since it aims at being fulfilled through its *non-recognition*, while the other two operate in a converse manner. The manipulative intention accompanies the ostensive stimulus, but it is not an intention to ostensibly *mean* something. Now, when a speaker is attempting to manipulate, one way of getting the addressee not to suspect the manipulative intent is to get him to process the information in a completely standard way, as it is crucial for the success of the manipulative attempt that the utterance actually meets the usual expectations the hearer has by virtue of the presumption of relevance, and so it is instrumental to the success of manipulation that whatever the speaker utters is processed by the audience as bearing both a communicative and an informative intention. To this extent, manipulative communication bears a level of genuinely standard communication: meaning gets through, and the way this meaning is inferred by the hearer is no different than the way meaning is inferred in standard, cooperative, conversational exchanges. Moreover, leaving aside considerations about the deceptive nature of manipulation, the communicative exchanges through which it operates do satisfy the expectations of relevance the hearer has.

The main problem we have to account for manipulation, thus, could be formulated as follows: we will not be able to distinguish between manipulative and non-manipulative communication using RT's postulate of double intentionality since both bear informative and communicative intentions the recognition of which is crucial to the success of communication. The difference is elsewhere, and the milk example (formulated in (69) above) showed that it seems to reside above all in the fact that an additional intention motivates the speaker, an intention which must remain concealed. Even if this problem of intentionality appears to be one of the most complex issues associated to manipulation, I will suggest that the difficulty can be dodged by looking at things from a different angle, since RT has more to offer than other pragmatic approaches. The main difficulty here presumably comes from the conception of the issue in terms of a tension between the 'what' question and the 'how' question. If we focus on the first, it seems that RT's conception of ostensive-inferential communication shares the same principle of overtness as SAT's and Grice's. Nonetheless, I will try to show in what follows that RT turns out to be an extremely useful framework to address the 'how' question, as I contend it can accommodate quite nicely the idea that manipulation consists in having the hearer reach an interpretation which he should rule out should he be aware of certain inconsistencies he is misled into missing.

Manipulation is about misleading the audience into inferring sets of assumptions which would yield (logical or pragmatic) inconsistencies if processed against an adequate context. We might find opportune to try to translate this idea with the help of the cognitive notion of relevance, as defined by Sperber & Wilson. Therefore, I suggest that we could interpret the manipulative

intention as the intention of having the content of the speaker's informative intention achieve a satisfactory degree of relevance (see definition (32) given in §I.3.2.2.); said more technically, the content of the manipulative stimulus must be such that the amount of cognitive efforts it requires to be derived will be compensated by the contextual (and cognitive) effects it will yield. But we easily see how this formulation applies in fact to all communicative stimuli (according to RT), and thus we need to specify it so as to make it discriminatory. I suggest that the second part of the operational definition of manipulation I gave in (32) makes for this need: manipulation operates by making the content of the informative intention relevant enough *so that* the hearer does not engage in further processing, which could tentatively lead him to uncover some inconsistencies between the content he derived and his cognitive environment, which in turn might lead him to suspect non-cooperative intentions on behalf of the speaker. This is the line of investigation explored below.

I will in particular develop the idea that RT provides an account of how meaning is derived, which should in principle take care of the first part of my operational definition of manipulation; it moreover offers a way of explaining why and how manipulated hearers come to miss certain inconsistent elements the moment we assume that this can be a consequence of the stimulus achieving a satisfactory degree of relevance, which is a first step into providing clues to resolve the second part of the definition. For the time being, let me assess the principled legitimacy of a framework such as RT for an explanatory account of the 'how' question that motivates scientific research. For this, I will take advantage of the fact that RT is more than a pragmatic theory, since it also makes more general claims about human cognition.

III.1.3.4. Multidisciplinary advantages of a cognitive pragmatic account of manipulation

Additional motivations can support the choice of a cognitive pragmatic model of information-processing to address the subject of manipulative communication. First of all, such a framework accommodates nicely with mainstream psychological and cognitive models that provide evidence for a conception of the human mind as a fallible information-processing device, such as Kahneman & Tversky's theory of rational decision (see §II.3.1.2.). As mentioned before, these heuristics may be taken to also characterise tendencies in language processing mechanisms. An additional psychological trend, perhaps broader in its scope, known as the study of cognitive illusions (cf. Pohl 2004 and section §II.3.1.3.), converges in this direction by showing how humans are often subjected to cognitive errors, and is meant to tackle "a plethora of phenomena showing that we deviate in our thinking, judgement, and memory from some objective and 'correct' standard" (Pohl 2004: 1). By relying on a perceptual analogy, research on cognitive

illusions considers that the mechanisms involved in thinking, managing memory and providing judgements are, as in the case of optical illusions, subject to errors.[205]

As RT's model of information-processing inherently incorporates the possibility of misunderstandings and errors in communication, I take it that it can in principle be interfaced with frameworks in cognitive psychology that address precisely phenomena related to errors or mistakes in judgement, memory or reasoning. In what regards the phenomenon of manipulation, I furthermore hypothesise that such an interface can prove extremely fruitful, to the extent that it makes sense to consider, following Rigotti, that "the dynamics of manipulation are very close to the dynamics of human error" (Rigotti 2005: 69). Many examples of successful manipulation hinge upon the speaker's expectations that the hearer will miss crucial information or fail to give it proper attention; furthermore, the literature on cognitive illusions (e.g. Pohl 2004) provides thorough studies of cognitive biases which can be used to explain certain instances of manipulation (see Maillat & Oswald 2009, forth.). As RT provides a model which mechanistically maps processes of interpretation, it has room for considering different types of constraints that may affect the interpretative process. I take cognitive biases and neighbouring cognitive phenomena to represent some of these *internal* constraints.[206]

A second advantage of adopting a cognitive pragmatic framework is that it can in principle also host and account for the role of *external* constraints, such as social processes which have been identified as playing a role in mechanisms of influence.[207] Processing information in peer pressure contexts, or when the information stems from a renowned expert on the topic at stake may definitely affect our attitude towards the information, in particular in what regards mechanisms of belief fixation, since such constraints may either bypass or slow critical vigilance. RT's model may incorporate these parameters by conceptualising them as contextual assumptions that will affect the degree of epistemicity we are able to assign to the utterances.

The originality of my proposal resides in the incorporation of an account of these constraints into the domain of communication. One of the fundamental assumptions of my research stipulates that these constraints, whether internal or external, have counterparts which are realised in and through language, and I intend to take advantage of the interfaceability, into a model of information-processing, of research in pragmatics, argumentation theory, linguistics,

---

[205] See §II.3.1.3.

[206] Internal here denotes properties of the cognitive system itself.

[207] See Cialdini (2009) for an informed review of research in psychosocial science and my review of some of the work in that paradigm in Chapter II §II.2.1.

discourse analysis, cognitive psychology and social psychology in order to show how these constraints may be manipulatively enforced and exploited in discourse.

A last word regarding the perspective I am speaking from: attempting to capture manipulation through a cognitive pragmatic model of communication such as RT allows looking at things from the perspective of the hearer, and this, I contend, may provide novel and scientifically challenging clues regarding the 'howness' of manipulation. As I have tried to show through my discussion of the complexities linked to the 'what' question, the intentional status of manipulation coupled with its covert nature calls for an examination of the manipulator's mental states (intentions, goals, beliefs, desires, etc.), which is an opaque area of inquiry, since those mental states are firstly private, and secondly seldom made public.[208] Adopting the perspective of the hearer instead gives us potential insights into the mechanisms by which the constraints evoked above can operate, especially when we assess their role in the processing of information. The detailed elaboration of the model I advocate will take place below.

III.2. COMPREHENSION AS THE OUTCOME OF A FALLIBLE PROCESS

What the manipulator does with words can be conceived as an analogue of what the illusionist, or magician, does with her/his hands: the latter waves one hand in front of you to catch your attention with it, while s/he is doing something with the other hand – something which you do not focus on and which is crucial to the success of the trick. One of the recognised skills of magicians is indeed their ability to draw your attention to what they want you to focus on, so that they can freely perform whatever they need to perform in view of succeeding in the trick (what could be referred to as the method) without their target being aware of it, as their attention is captured by something else: "Magicians are, first and foremost, artists of attention and awareness. They manipulate the focus and intensity of human attention, controlling, at any given instant, what we are aware of and what we are not." (Martinez-Conde & Macknik 2008: 46).This is presumably also something that skilled pickpockets are good at doing, because it is crucial for them that your attention is drawn elsewhere than on the pocket they are aiming for and on the method they use to steal something from you.[209] The idea I will develop here is that similar patterns of attention misdirection can be identified in manipulative communication.

---

[208] I assume following van Dijk (2006) that few manipulators would admit to their manipulative attempts.

[209] Actual footage of pickpockets at work on French TV revealed for instance that they distract their target's attention by asking directions and trying to get them to focus on a rather large map of the city they unfold in front of them. This allows them to conceal what is happening underneath the unfolded map (i.e., the pickpocket searching her/his victim's pockets).

Manipulation of attention, for magicians or pickpockets, is meant to ensure two things: i) that the targets only process the (perceptive) stimuli that the trickster wants them to process, which are the stimuli that are not detrimental to the success of the trick, and ii) that the targets do not process other stimuli that, if made salient, would make the trickery collapse altogether, such as for instance the precise movement where the magician or pickpocket takes the wallet out of the pocket of the target. There are obvious similarities between this definition and the definition of manipulation I gave in the first chapter of the present work (see §I.3.2.) if we put on a par processing of perceptual information in the case of magic and processing of communicated information in the case of manipulative communication: let me recall that manipulation, in mechanical terms, is a twofold process whereby some information is made salient while other is obscured. Since information directed at someone can be conceived as verbal stimuli, my definition of manipulation can be rephrased as a twofold process whereby some stimuli are made salient while others are obscured. Here we find exactly the principle with which magicians and pickpockets operate. The hypothesis I therefore want to develop now is that, pretty much in the same way as we can start explaining how attention is directed towards certain stimuli and driven away from other stimuli,[210] there are ways to account for how information is likely to be selected as input in the meaning derivation procedure (and thus be made available for further processing if need be), but also for how information is likely not to be selected as input in the meaning derivation procedure. To this effect, I will assume that the notion of cognitive relevance, as defined in Sperber & Wilson (1995), is suited to explain how manipulated addressee's attention can be constrained.

III.2.1. RELEVANCE DOES NOT GUARANTEE COMPREHENSION

III.2.1.1. Relevance, processing effort and cognitive effect

In RT's model of communication, it is assumed that the comprehension procedure is triggered by the hearer's recognition of the speaker's communicative intention, the latter being defined as the speaker's making it mutually manifest that she intends to convey a certain set of assumptions to the hearer. In other words, ostensive stimuli carry a *presumption of relevance*. By making (mutually) manifest her intention of making manifest a set of assumptions to the hearer, the speaker *ipso facto* "necessarily communicates that the stimulus she uses is relevant to her audience" (Sperber & Wilson 1995: 156). This means that whenever a hearer notices that an

[210] Research on attention focusing has shown for instance how certain types of movements can attract more attention than other movements (e.g. elliptical and round arm movements instead of straight and short movements, see Martinez-Conde & Macknik 2008); I hypothesise that similar dynamics underlie manipulative communication on the speech level.

utterance is directed at him, he will assume that the content of the utterance is worthwhile processing, for otherwise the speaker would have no reason to produce an ostensive utterance directed at him. Such a recognition functions as a trigger for the hearer's inferential processing of the utterance. Otherwise known as the recognition of the presumption of relevance, this phenomenon, in order to be completed – that is, for comprehension to take place and therefore for communication to succeed –, requires that the hearer will pick up the set of assumptions the speaker intends him to process.

Our intuitive notion of relevance is developed in RT in technical terms as a relation between an utterance and a set of assumptions (or context); moreover, it is a relation endowed with a convenient – in terms of efficiency – ratio between processing effort and cognitive effect. Assuming that human cognition is an efficient information-processing device, Sperber and Wilson also claim that it is

> "aimed at improving the individual's knowledge of the world. This means adding more information, information that is more accurate, more easily retrievable, and more developed in areas of greater concern to the individual. […] Thus, long-term cognitive efficiency consists in improving one's knowledge of the world as much as possible given the available resources." (Sperber & Wilson 1995: 47)

Under this perspective, the relevance of an utterance depends on the ratio between processing effort and cognitive effect involved in its interpretation (which itself determines the overall efficiency of the utterance as a means to convey speaker meaning): focusing on the different conditions exposed in the above quote, we can say of an utterance that it is relevant when it adds more knowledge to the individual's cognitive environment (defined as "the set of assumptions available to him" (Sperber & Wilson 1995: 47)), when it provides accurate information, when it provides information that is useful to the individual's goals, but also when this is done to a low processing cost. Consequently, relevant utterances will be those which require little cognitive processing effort while at the same time yielding greater cognitive effects, the latter being defined as an enhancement of the individual's knowledge (of the world).[211]

I have said in the preceding section (see §III.0.) that one of the ways of constraining an interpretation in a manipulative attempt is to constrain the input of cognitive calculations. I will thus have to specify this claim in order to describe and explain, in light of RT's account of communication, in which ways inputs to cognitive processes can be constrained.

---

[211] Cognitive (or contextual) effects are defined in RT either as the addition of new information, the revision or the deletion of previously held assumptions.

The advantage of such a cognitive approach to communication is that it licenses a characterisation of relevant utterances by listing their properties in terms of processing variables, such as effort and effect. This has an important methodological consequence: it allows us to hypothesise that variation in the parameters (processing effort and cognitive effect) will result in variation in relevance; in other words, constraints on processing effort and on cognitive effect will result in constraints on relevance. Recall that manipulation is about making sure that certain information is represented and other information is left out of the picture, which could be reinterpreted as a constraint of relevance, for irrelevant information will tentatively be left out, while relevant information will presumably be represented. With RT's account of comprehension, we thus have a way of accounting for how this is possible by taking a closer look at the ways in which speakers may be able to set constraints on hearers' interpretations by manipulating the relevance of their own productions, i.e., by manipulating the processing effort and cognitive effect their utterances are likely to respectively require and yield in the comprehension process.

III.2.1.2. Context

A last but crucial aspect of RT should be dealt with before laying out the proper cognitive account of manipulative communication I am advocating here; I have exposed the type of constraints that can be set upon processing variables, but I have not talked about what these constraints bear on, in representational terms.

When we process verbal stimuli, we do so against a set of assumptions about the world called a context. Contexts are the result of a selection our minds operate on the totality of our encyclopaedic knowledge but also on perceptual assumptions that are immediately accessible, such as the space, time and physical settings of the communicative exchange, and on assumptions which the individual is able to infer from available information. Speber & Wilson (1995: 39) call this last entity an individual's *cognitive environment*, also defined as the set of assumptions that are manifest (i.e., perceptible or inferable) to an individual; contexts are therefore conceived as subsets of the cognitive environment. It follows that contexts are not given, but constructed,[212] since they result from a selection of assumptions and change constantly as information is gradually added and/or updated in the course of communicative exchanges.

[212] I use the term 'constructed' here to denote the fact that contexts result from a cognitive operation, not in the constructionist sense of construction. Many thanks to Patrick Morency for pointing out the ambiguity.

Context formation follows a dynamics of relevance: the assumptions that will make it to the context are those which are relevant to the task at hand, i.e. in what interests us here, mostly comprehension. This means that the assumptions that will make it into the context are those that are relevant; in other words, those that require little processing effort and which yield significant cognitive effects.

RT regards contexts as gradually-constituted cognitive entities: the context will start off with immediately accessible or perceptible assumptions (such as encyclopaedic entries and perceptual data) and will add assumptions as it sees fit, i.e., as long a relevant context of interpretation has not been reached. Let me take an example to illustrate:

(72) "Peter: Do you want some coffee?

Mary: Coffee would keep me awake." (taken from Sperber & Wilson 1995: 34)

In order to understand whether Mary's implicit answer amounts to a 'yes' or a 'no', Peter must first mobilise general assumptions he already entertains about the world, and which are relevant with respect to the content that is explicitly encoded in Mary's answer. In this case, relevant information would be information Peter has about the relationship between coffee and sleep, so information such as 'coffee is an excitant', 'people who want to stay awake are likely to drink coffee', or 'people who want to sleep are likely not to drink coffee', which may serve him to draw further consequences. Then Peter combines to this set of assumptions information which is immediately accessible from his environment, such as 'Mary seems to be enjoying the evening so far' or 'Mary is yawning and looks like she wants to go home to rest', depending on what Peter can tell from Mary's behaviour. Combining this set of contextual assumptions to the utterance – which has given Peter a sense of which relevant assumptions to look for – Peter will be able to derive an interpretation ('yes' or 'no'). In the 'no' scenario, this would be formalised in a deduction-like manner, such as in (73):

(73) a. People who do not want to stay awake are likely not to drink coffee

b. Mary looks like she is tired and wants to go to bed

c. Therefore, Mary's answer means that she does not want coffee

Crucially, for our purposes, RT assumes that the incremental construction of contexts has a psychological counterpart, whereby "order of inclusion corresponds to order of accessibility" (Sperber & Wilson 1995: 142): the first assumptions that will make it into the context are the most accessible and salient ones, i.e., those that require little processing effort on behalf of the subject. Conversely, the more efforts it takes to access a context, the less relevant the context will be.

RT's construal of the notion of context relies on the idea that all information that is (potentially) relevant in a given conversational exchange is not guaranteed to make it into the context, and this means that not all information is equally accessible. This incidentally explains the very possibility of misunderstandings. I recently found an example of someone intuitively describing this very idea in an internet forum devoted to cultural misunderstandings. The person who posted the comment wrote:

(74) "Vietnamese people tend to ask about your age the first time they meet you. It is because Vietnamese people use different words to call you in conversations, depending on your age, your social position, your relation with other people around them."[213]

Imagine that in a conversation a Vietnamese man asks an American woman for her age. She might take it as an offence, as in Western societies asking someone's age, in particular a woman, is considered to be a *faux pas*, and end up understanding that her interlocutor is trying to make a pass at her. However, what he means is literally to know her age in order to have more information about her, so that he can address her properly in the future, according to his own politeness standards. The misunderstanding could have been avoided should the woman have been able to mobilise the assumption that in the Vietnamese culture asking for someone's age is meant to provide information related to practical politeness issues, or if the Vietnamese man had been able, before asking his question, to represent the fact that asking a woman's age is usually considered to be rude in Western societies. With this example we see that failure to mobilise one relevant assumption in the context can jeopardise the output of comprehension and yield misunderstandings. Such assumptions are only culturally salient in the example (and thus it is for cultural reasons that alternative assumptions about politeness become less available to a given community).

Example (74) shows that inaccessibility to relevant assumptions can be responsible for misunderstandings. As I have already noted, there are grounds to contend that the dynamics of misunderstanding resemble the dynamics of manipulation; one way of accounting for that is to assume that both phenomena are genuine instances of communication, albeit failed ones. Both involve the same cognitive processes, which are standard comprehension procedures, and we will see further along that standard mechanisms involved in the selection of contextual assumptions may account for both (see §III.5.2.).

[213] Example found on http://www.english-test.net/forum/ftopic11240.html, last accessed 09.08.2010

III.2.2. FALLIBILITY OF THE COMPREHENSION PROCEDURE

As I have tried to show in this section, reaching an accurate interpretation of what a speaker means is not an easy task; strict identity between speaker meaning and the hearer's interpretation can only be speculated about, and all one can reasonably hope for is that the interpretation arrived at sufficiently resembles speaker meaning (cf. Wilson 2000). Hence, the comprehension procedure is fallible, and it should be noted that this is the case at the level of both conceptual entries and contextual information management.

Processing communicated concepts is a cognitive operation which affords significant interpretative latitude, to the extent that their meaning can vary greatly, either through narrowing or strengthening operations: as noted by Carston,

> "a well-established feature of the relevance-theoretic view of verbal comprehension is that concepts lexically encoded in the linguistic expression uttered may be pragmatically strengthened (or narrowed) as part of the process of deriving the intended explicit content of the utterance." (Carston 2002b: 83)

Ad hoc concept formation is the operation by which a selection of the relevant lexical properties of a concept is performed in an attempt to comprehend what the speaker meant when she uses that particular concept in a communicative exchange, and in particular when it is not used in its standard acceptation. For instance, when saying that "Sally is looking for a bachelor", granted that it is contextually salient that Sally wants to settle down and have a family, then what the speaker probably means by BACHELOR* is a subset of males satisfying the condition of being eligible for marriage, not any one from the whole range of individuals technically denoted by the concept *bachelor*, which includes 80 year-old widowers or teenagers.[214]

Now, the interesting feature of the notion of *ad hoc concept* developed within RT to characterise the way we process lexical entries on the basis of contextualised linguistic stimuli is that it inherently carries the possibility of misunderstandings arising out of a mismatch between the hearer's actual and the speaker's intended selection of relevant lexical properties. Since a significant part of the interpretative work is the hearer's responsibility, there is always be the possibility that speaker meaning will not in fact end up sufficiently resembling the interpretation derived by the addressee; moreover, this is something that manipulative communication could exploit, because it creates the possibility of using concepts in a very narrow or on the contrary loose way, so that their referring potential varies accordingly. I have already mentioned (see §II.3.1.3.) that Allott (2005) and Allott & Rubio Fernández (2002) showed that this can happen

[214] The standard notion of *ad hoc* concepts in RT is to write the word in capital letters with a star to specify that it is to be interpreted in an *ad hoc* way.

in political discourse for keywords such as *democracy*, sometimes taken to mean something like "freedom of investment". The same *ad hoc* concept, DEMOCRACY*, is oftentimes explicitly summoned in heated discussions where a participant, tired of being interrupted, demands the right to be heard by invoking democracy ("Aren't we in a democracy? Everyone has the right to speak!"): democracy is certainly partly about plurality and freedom of speech, but the use of the concept is such situations is presumably only meant to exploit the positive connotations of the term, and not its atomic denotation. The exploitation of a*d hoc* concepts can thus represent a valuable strategy in view of furthering argumentative, persuasive but also manipulative goals.

I have also defended this assumption on a contextual level by considering how addressees may fail to mobilise the appropriate contextual assumptions in the comprehension procedure. This is a natural property of the cognitive principles governing utterance interpretation: we may perfectly find the relevant assumptions and get to what the speaker meant, but since we are responsible for the mobilisation of contextual assumptions, sometimes we may fail in calling in the correct (intended) assumptions and thus fail to reach the appropriate result. In other words, it is precisely because the hearer is responsible for the derivation of certain contextual assumptions that speakers cannot hope for a 100% success rate in comprehension. The only way to secure such a failsafe result would be to utter exclusively explicit contents in an attempt to control all information that will be taken into account by the hearer. But this is not how communication works, if only because of time and resource constraints; what speakers do as they produce an utterance amounts to facilitating the hearer's cognitive processing by laying out manifest cues that are instrumental to deriving the intended interpretation and that will hopefully be picked up, not to providing all and every information necessary for the interpretative mechanisms to yield the correct output. This represents both a cognitive limitation and a significant advantage of communication; on the one hand it allows maximising the efficiency of resources without overly compromising the reliability of the output, even if on the other hand it does sacrifice 100% reliability.

Combined with the arguments concerning the heuristic nature of comprehension, these observations allow me to conclude my exposition of the aspects of RT that I think are relevant to an account of manipulative communication. As way of summary, let me list some assumptions about communication which stem from the adoption of RT as a framework and that I will retain as crucial:

- The processes governing utterance interpretation are heuristic: they are fast, efficient, automatic, and fallible in that they do not rely on systematic processing of the input.
- Variations in processing variables such as cognitive processing effort and cognitive effect result in variations in the relevance of the processed representations.

- Consequently, there is room to explore more in depth, as far as manipulative communication is concerned, if there are ways to manipulate the relevance of input representations to the comprehension process.
- Since the construction of context obeys the relevance dynamics, it cannot constitute a failsafe procedure and may generate misunderstandings, or more generally interpretive problems (with respect to all potentially relevant information that an addressee could take into account). Thus, the same mechanisms of relevance manipulation that apply for communicated representations also apply to contextual assumptions.

Relevance Theory is a framework meant to explain the cognitive underpinnings of communication, and in particular to uncover the conditions under which communication is successful. The model defended in RT is also especially suited to give an account of misunderstandings and other types of errors that one is prone to making in communicative practices.

This is precisely the reason why RT constitutes such an appealing framework to account for manipulative communication. Rigotti argued that "the dynamics of manipulation are very close to the dynamics of human error" (Rigotti 2005: 69); I will try to show in the following sections that the mechanisms underlying manipulative communication are actually the mechanisms underlying communication *tout court*, and that the success of manipulative attempts can be explained with the same tools that explain both the success and failure of standard communication. This is a stance that explicitly takes the hearer's perspective: in contrast to alternative perspectives, from which manipulation is in my opinion problematically accounted for (see §I.3.1.), a cognitive pragmatic perspective focuses not on the properties of manipulation as a phenomenon, but on the cognitive operations that manipulation prompts for in its targets. As already noted (see §I.3.2.), this carries interesting consequences: it can get rid of the problems associated to second-hand manipulation, but also puts on a par *the effect* of manipulative communication with that of fallacious argumentation, even if the latter is not necessarily manipulative. I will contend (see §III.5.2. and §III.5.3.) that this, is of particular interest to theories of argumentation.

III.3. THE CONTEXTUAL SELECTION CONSTRAINT (CSC)

III.3.1. GENERAL OVERVIEW

RT's account of comprehension elicits the crucial idea that not all relevant pieces of information are *ipso facto* guaranteed to make it into the context of interpretation; I contend that this is precisely what manipulative utterances exploit and now turn to presenting the Contextual Selection Constraint (henceforth CSC), which is to a large extent the fruit of a research collaboration with Didier Maillat (see Maillat & Oswald 2009, forth).

Postulating that manipulation operates a constraint on information-processing requires that we specify the scope of the constraint. I have ventured at the beginning of this chapter (see §III.0.) that a manipulator may constrain the input of the comprehension procedure but also the type of cognitive processing itself (i.e. systematic *vs*. heuristic processing). I will now focus my attention on the first object of constraint, namely the communicative input of the comprehension procedure.

I refer to communicative input as the set of assumptions that the addressee will take into account as he works out speaker meaning, or the set of assumptions that are made mutually manifest. Manifest inputs can be further broken down into two categories: assumptions which are contained in the message (i.e. explicit information which is directly accessible), and contextual assumptions.[215] In order to constrain comprehension, a manipulator may hence either constrain the explicit content of her message or try to constrain its context of interpretation.

Constraining the message itself amounts to controlling the explicit information you choose to encode and the linguistic material you choose to communicate it with, but this is rather trivial and of limited interest for a comprehensive account of manipulation for at least two reasons. First, there are no 'formal' traces of manipulation in that the same utterance may be manipulative in certain cases and not in others (see the 'milk examples' (27) and (69));[216] even if we were to identify in a given discourse several rhetorical strategies such as the use of tailor-

[215] To some extent, this bipartition echoes Sperber and Wilson's definition of manifestness: "to be manifest is to be perceptible or inferable" (Sperber & Wilson 1995: 39). In this case, assumptions contained in the message could be categorised as (communicatively or explicitly) perceptible, while contextual assumptions would be defined as inferential, to the extent that their mobilisation is the result of inferential work on behalf of the addressee.

[216] (27): "Laszlo, there is no milk left for tomorrow's breakfast; could you go to the grocery store and buy some more?"

(69): Could you go to the grocer's and get a few cartons of milk? We've ran out of milk and it would be nice to have some for tomorrow's breakfast.

made metaphors for instance (see Oswald & Rihs 2010),[217] it would still not mean that they are characteristic of manipulative communication – and there are in fact no such thing as stable linguistic forms that would be specific to manipulative discourse.[218] Additionally, limiting the analysis to the identification of specific formal or linguistic devices that may be used to further manipulative goals might fall short of providing a proper explanation of manipulation; if certain types of expressions have been shown to be instrumental to the success of a manipulative attempt, it is not solely by virtue of their linguistic properties, but indeed mainly because they trigger specific cognitive operations which will mislead the addressee, and this is where the bulk of the explanation – and thus the scientific interest – of a pragmatic account rests. Second, the explicit content may be untruthful, unreliable or uncooperative in Grice's sense, but this is not a specific condition of manipulation either – and furthermore this is not something that will help us decide whether the utterance is manipulative or not (see §I.3.1.1.1.).

More generally, exclusive focus on constraints on the message itself (i.e. constraints on its explicit content and on its form) are of limited interest for the type of account I am trying to develop if they are not systematically linked to the type of cognitive response they induce in their targets. Studies of production who dispose of reception are thus problematic insofar as they usually present discursive strategies without going beyond the simplistic assumption that these strategies are effective either because i) they rely on implicit information ('implicit' intuitively equals 'covert' and therefore prompts for suspicion), and ii) they implicitly assume that people's readiness to trust that their interlocutors will provide truthful information is a sufficient condition for the success of manipulation. This is not to say that a survey of the formal and semantic properties of verbal stimuli cannot help us to get a feel for some linguistic aspects of manipulative communication, but it does mean that as long as they are not combined with a thorough account of reception, studies of production fall short of psychological plausibility.

Yet, once we begin to bridge this gap by attempting to make systematic links between linguistic forms or discursive strategies and their cognitive effects, the study of input to cognitive processing becomes more likely to generate findings of great interest.

While the study of so-called 'manipulative' properties of utterances may seem ill-suited to fulfil explanatory requirements of a naturalistic approach to manipulation, such is not the case of the

---

[217] In this paper we argued that specific conceptual operations such as metaphors, which are realised linguistically, could yield perlocutionary effects such as boosting the credibility of the speaker and reinforcing the feeling of belongingness to a same community of interest.

[218] If they existed, we would be faced with the paradoxical situation in which we can identify successful manipulative forms whose manipulative nature necessitates that they remain concealed.

study of the way contextual assumptions are managed in manipulative communication. Because by nature contextual assumptions are not part of the linguistic stimulus, an account of how they are selected, derived and put to use in the comprehension procedure will in principle shed light on some of the cognitive processes underlying information-processing. Thus, taking into account i) that contexts are conceived as sets of assumptions, and ii) that according to the operational definition of manipulation I gave (see §I.3.2.2.) the success of manipulation rests on the efficiency of the twofold process through which the hearer is led to process certain information while failing to process other information that would be otherwise relevant to defeat the manipulative attempt, I argue (following Maillat & Oswald 2009: 362-363) that manipulation can be interpreted in technical terms as a contextual constraint. I am now able to specify my principled definition of manipulation accordingly:

(75) Manipulation is a twofold process ensuring i) that the communicative stimulus is processed into a limited 'manipulation-friendly' context and ii) that the context is not expanded so as to avoid the inclusion of 'manipulation-unfriendly' or problematic assumptions.

Quite expectedly, the magician analogy comes to mind once again: the task of the manipulator is to ensure that only certain contextual information is accessed and at the same time to prevent the addressee from processing other contextual information. If the illusionist tries to get people to focus on certain gestures and not on others, the manipulator tries to get people to focus on certain contexts and not on others. In terms of information-processing mechanisms of comprehension, this is equivalent to a *contextual constraint* which manipulates people's attention and focus on specific sets of assumptions.

Within the framework of the contextual constraint, I will assume that a cognitive pragmatic analysis of manipulative communication has to take into account two different potential contexts of interpretation: the first is the constrained or limited context, which contains only those assumptions which are compatible with the success of the manipulative attempt – i.e., it is devoid of any of the problematic assumptions that would put the addressee on the track of detecting the manipulative intention, either directly or indirectly through further inferential work. Let us call this context *C* (following Maillat & Oswald 2009: 363). The second context of interpretation which is relevant to an analysis of manipulative communication is the context C augmented with precisely the latter assumptions – i.e., the context which prompts for the identification of internal or external inconsistencies and of anything that would potentially hinder the manipulative attempt. Let us call this context *C'*. Manipulation, thus, is about constraining the addressee's interpretation so that the interpretative procedure stops at C and does not proceed to C'. In terms of epistemic vigilance, this amounts to saying that the

addressee's screening mechanisms are misled into validating C as the relevant context without even suspecting that there could be an alternative yet relevant C'. An additional way to look at the phenomenon would be to argue that some of the assumptions contained in C' would clash with some of the assumptions of C, and that this is sufficient to generate a cognitive dissonance in the addressee's mind; in an attempt to resolve it, the individual may subsequently be led to suspect the presence of a manipulative intention behind the identified inconsistencies, this in turn constituting a first move towards the dismantling of the manipulative attempt. Therefore, we can also consider that the success of manipulation requires that the potential cognitive dissonance between C and C' be concealed.

Manipulation can thus foremost be conceived as a constraint on the message's context of interpretation; this means that the model will have to account for how certain assumptions of the addressee's cognitive environment end up selected to partake in the calculation of speaker meaning, while others are likely to remain masked. This naturally subdivides the explanatory task in two: first, I will need to characterise how a manipulator can constrain the interpretation within a limited context C. To explain this phenomenon, I will assume that manipulators try to make the target contextual assumptions become cognitively inescapable by boosting their contextual relevance – i.e. by making sure that they meet the expectations of relevance generated by the speaker's utterance and the broader communicative situation. In this sense, manipulation is about strengthening certain assumptions. Second, I will assess the negative counterpart of this first process by taking a look at specific strategies that might block access to the alternative context C' which contains assumptions that are detrimental to the manipulative attempt, or which could serve as a basis to infer further assumptions that would be detrimental to the manipulative attempt. In this sense, the manipulative attempt also builds on a weakening strategy.

The explanatory task is therefore divided in two, given that we need to account for how both C and C' are constrained, the success of manipulation requiring that the former be fostered and the latter blocked. As my exposition of these strengthening and weakening strategies unfolds, we will see that the relevant parameters should not be taken into isolation from one another: by definition, fostering C implies blocking C' and vice-versa to the extent that I characterised manipulation as a twofold process, and thus I should make it clear that the involved strategies are not mutually exclusive. It could hardly be otherwise in an RT-based framework: relevance is a relation between an assumption and a context, and furthermore it supposes a ratio between processing effort and cognitive effect, which are the two information-processing variables on which I take manipulative strategies to operate. For ease of presentation and clarity reasons, I

will describe them one after another, but it should be borne in mind that by no means this supports the view that these strategies are autonomous.

III.3.2. SALIENCE AND RELEVANCE

We have seen that according to RT, relevance is achieved when processing effort is kept to a minimum and cognitive (or contextual) effects are important. In relevance-theoretic terms, relevance is achieved by meeting two *extent conditions*:

(76) "Extent condition 1: an assumption is relevant in a context to the extent that its contextual effects in this context are large.

(77) Extent condition 2: an assumption is relevant in a context to the extent that the effort required to process it in this context is small." (Sperber & Wilson 1995: 125)

Therefore, one way of making sure that a target assumption will be processed is to render it relevant, that is, to try to fulfil its extent conditions, either by lowering its cognitive processing cost or by making sure that the assumption contributes significant cognitive effects.

The same applies for sets of assumptions, i.e. contexts. The less processing effort contexts will require, the more they will be relevant, and the more cognitive effects their combination with a given utterance will yield, the more relevant they will be. With respect to the success of manipulative communication, this means that context C should easily meet both extent conditions, and that context C' should not.

I would like to introduce the notion of salience to help characterising the CSC, as I believe that the two extent conditions of relevance can both be interpreted as constraints on salience. Sperber and Wilson defined relevant information as "information that is more accurate, more easily retrievable, and more developed in areas of greater concern to the individual" (Sperber & Wilson 1995: 47). From this quote, we can extract three properties relevant information should have: first, it should be more accurate, that is, strong in terms of reliable information, or information that can safely be taken to correspond to actual states of affairs; second, it should be easily retrievable, that is, somehow made accessible in the context of the communicative exchange; third, it should be more developed in areas of greater concern to the individual, meaning that it should strongly be related to the individual's goals, in terms of its tentative instrumentality to the fulfilment of those goals. The first and third properties can thus be said to relate to extent condition of relevance #1, while the second answers extent condition of relevance #2. I also contend that these properties may be interpreted as bearing on the salience of information.

If an assumption is accurate and represents a very likely state of affairs, we can say it is salient in that it will stand out as a preferred candidate among other less reliable assumptions; if an assumption is accessible in a given communicative exchange, it is *ipso facto* salient compared to other less accessible information; if an assumption is more developed in areas of great concern to the individual, we are also entitled to consider the assumption to be salient because its instrumentality to fulfil some of the individual's goals will make it stand out among other less useful assumptions. These three parameters in principle should make for the main condition of salience, namely that a salient item is one that is "foremost in someone's mind" (Giora 2003: 15). Salience is a comparative notion, as by definition it denotes the property of one object which it precisely does not share with the other objects around it: it is in the very essence of salience that it expresses a property which other comparable items do not have. In this sense and given my twofold definition of manipulation, salience is theoretically compatible with the account I am after, since I assume that the success of manipulation rests both on the representation of a context C and on the non-representation of context C'. I consequently postulate that context C is made of salient assumptions and that C' is made on non-salient assumptions.

According to Giora, who develops a pragmatic account of meaning centred on the meaning of linguistic units – thereby reducing the importance of contextual factors –, salience can be defined in terms of frequency, familiarity, conventionality and prototypicality. Note that frequency, familiarity and conventionality are elements which resonate with the accessibility extent condition, while prototypicality may be taken to relate to the strength extent condition. Thus, for an assumption to fulfil frequency, familiarity, conventionality and/or prototypicality requirements in a given community of language users amounts to the assumption being salient within that community. With respect to information-processing mechanisms and bearing in mind the incremental construction of contexts, salience is a condition that will be instrumental for some assumptions achieving relevance (as opposed to others); in this sense, the relationship I see between salience and relevance is one of facilitation: the more a speaker tries to endow certain assumptions with salience in a given communicative exchange, the more chances they have of achieving relevance by virtue of the effect salience has on both extent conditions of relevance.[219] Put differently, since salience facilitates assumption accessibility and fosters assumption strength, salient assumptions will tentatively have more chances of being taken as input for the comprehension procedure than non-salient ones, for the latter require the allocation of more cognitive resources to be derived by the addressee. This applies both to explicit and contextual assumptions: explicit contents are presumably inherently salient, since

[219] But this does not mean that salience is a sufficient condition for relevance.

they are directly presented to the addressee's attention and contextual assumptions may be rendered salient by tweaking information-processing variables such as accessibility and strength.

One further interesting aspect of the notion of salience is that it accommodates rather nicely with some cognitive biases exposed previously, namely the mere exposure effect and the validity effect (see §II.3.1.3.2.). The mere exposure effect generates judgements of liking, while the validity effect generates judgements of validity; both types of judgements, in my view, can be interpreted as having an effect on the salience of information. Familiarity, triggered by repeated exposure, can contribute to liking via the mere exposure effect; more generally, we can extrapolate that since repetition of certain information fosters favourable attitudes towards the repeatedly accessed object, it *ipso facto* fosters its salience, for usually things we like stand out compared to things we have no particular attitudes towards. Hence, the more someone experiences and represents an object, the more the object will readily pop to mind provided circumstances permit it (i.e., the more the object will be salient), to the extent that its having been processed repeatedly in the past will presumably decrease the efforts needed to process it again. On the strength side, the more certain information is repeated, the more valid it will be perceived to be (following the validity effect): repetition increases strength of assumptions, and strong assumptions are salient compared to weak ones, making them preferred candidates for input to further cognitive processes.

The example of advertising slogans illustrates this vividly: they work as efficient commercial strategies because they are presumably catchy, easy to remember and most importantly repeated as much as possible. Attention-grabbers such as slogans, in this sense, play a specific role in trying to increase the chance that the input they are associated with will actually serve as input when decision-making is involved, for instance when someone needs to buy a specific product: the ones that have been advertised repeatedly in the past will tentatively be preferred thanks to their ease-of-processing property and salience, which is a first indication to understand why they are the first that come to mind when envisaging the category of products involved. On a similar line of argument, the interpretation of salience both in terms of strength and accessibility also plays a part in the explanation of why speakers who wish to convince an audience might want to try to appeal to the values that seem to be shared by the members of their audience: these values are deeply embedded in their cognitive environment and thus they are salient pieces of information on which they base their behaviour; any piece of information related to them will presumably automatically trigger their activation and subsequently achieve relevance.

Salience is a property of utterances which straightforwardly accommodates the economical requirements of information-processing postulated by relevance theory because of its influence on both processing effort and cognitive effects. The more salient the representation, the less effort it will take to be processed and the more it will carry cognitive effects; conversely the less salient the representation, the less likely it will be to partake in cognitive processes (which means that representing it will require more processing effort). This phenomenon is captured by Giora in the following quote: "while salient information is highly accessible, nonsalient information requires strongly supportive contextual information to become as accessible as salient information." (Giora 2003: 15).

To sum up, I will consider that salience is something that can affect the relevance of information in communicative exchanges, as salience is instrumental to the fulfilment of extent conditions of relevance #1 and #2. That is to say that salient information is at the same time *accessible* and *strong* information, this in turn making it hierarchically more eligible for inclusion as informational input to cognitive processing by virtue of its compliance with information-processing variables. In order to develop a cognitive pragmatic account of manipulation, hence, what I will need to do is to assess the different strategies that may render the preferred context C relevant to the individual and the problematic context C' irrelevant.

As briefly mentioned above, (see §III.3.1.), the characterisation of these strategies will not be an easy task to the extent that the type of constraining that one can impose onto the parameters influencing salience and relevance (i.e. onto processing effort and cognitive effect) are two sides to a same coin when considered in cases of manipulative communication: since manipulation is a twofold process, trying to render context C salient will automatically result in making C' less salient, and vice-versa. This means that there will be 'context-strengthening' strategies as well as 'context-weakening' strategies (cf. Maillat & Oswald 2009, forth.), but that both will inevitably have mutual implications: if an assumption's salience is increased, this means that another assumption's salience is made lower, and this is exactly what happens in manipulation according to my twofold definition. In the following section I will therefore expose several of these strategies by trying to make explicit which context they are aimed at constraining (whether they boost the salience of C or lower that of C') and at the same time I will try to state the effect each has on its contextual counterpart.

III.4. Operational aspects of the CSC

Since manipulation is defined as a twofold process, the Contextual Selection Constraint model used to account for it will need to explain how constraints operate both on context C and on context C'. In the first part of this sub-section I will focus on the strategies that manipulators may

employ to directly constrain the input to cognitive processing, so that it remains limited, i.e., devoid of any assumptions that would be detrimental to the manipulative attempt. I will nonetheless highlight that these strategies have obvious indirect implications on context C', as, in light of the relevance context C has achieved thanks to the CSC, its chances of being processed will presumably automatically drop. In the second part of the section, I will focus on the strategies that are specifically directed at preventing context C' from being processed and found relevant: these include mechanisms of salience reduction, but also bear indirect implications on the salience of context C, to the extent that whatever was tentatively considered for inclusion in C', as a result of the CSC, has been deemed irrelevant by the addressee, thereby leaving the door open for inclusion of other information in context C.

III.4.1. STRENGTHENING STRATEGIES: INCREASING THE SALIENCE OF CONTEXT C

Relevant information, in RT's terms, is information which, among other conditions, requires little effort to be processed. Therefore, I take it that one way of constraining a context is to ensure that the assumptions you want to include in that context require little processing effort, so that at least one of the extent conditions of relevance is fulfilled. Relevance can consequently (partly) be conceived as a function of information availability – and in particular of contextual information availability.

Assumptions requiring little processing effort are those which are readily or easily available to be picked up in the conversational exchange. Availability can thus be described in terms of salience, itself describable in terms of accessibility and strength. Following the path of least effort, the more accessible and/or the stronger the assumption, the more likely it is to make it into the context of interpretation (in this sense, it could be said that the most accessible of all assumptions are those that are explicitly formulated). A first parameter a manipulator may want to exploit is therefore extent condition #2 by increasing the accessibility of the assumptions she intends to enter context C. Let me look at some strategies aimed at achieving this.

III.4.1.1. Fostering accessibility of assumptions in context C

III.4.1.1.1. Repetition implements the CSC over time

Fostering the accessibility of assumptions can be achieved in several different ways; moreover, it is not a phenomenon that is limited to one single speech event. Alongside the mere utterance of explicit contents, which are trivially accessible because they are just there out in the open ready to be picked up, manipulators can for instance try to exploit the addressee's repeated exposition to a same stimulus in order to take advantage of its availability. This is where work on the validity effect (see Hackett Renner 2004 and §II.3.1.3.2.) might prove useful. The validity effect

stipulates that the more an individual is exposed to a stimulus, the more validity this stimulus will be endowed with in the individual's mind. I contend that this bias, which has experimentally been shown to be robust, plays an important role in mechanisms of conviction because it can enforce the combination of new information with previously known information retrieved in the form of relevant contexts. Let me take an example.

III.4.1.1.1.1. The 'Stauffer Case'

In October 2009, Eric Stauffer, leader of the populist Genevan political party *Mouvement Citoyen Genevois* (MCG), was running for a seat in the Genevan Executive Council. Throughout the campaign, he benefitted from high media exposure – as did the other candidates – to defend his ideas; he was mostly campaigning against French borderers, blaming the economic crisis and the high unemployment rate on them by claiming that they were stealing the jobs of the Genevan people and concluding that the Genevan policy regarding work permits for foreigners was overly generous. He usually backed up this standpoint with statistical figures he repeatedly presented as evidence for his claim. On November 3rd, the local newspaper, the *Tribune de Genève*, published an article disconfirming the numbers given by Eric Stauffer in his campaign:[220] Stauffer had claimed that the Genevan Public Transportation service and the Industrial services (electricity, water, gas) employed 50% of French borderers; after having thoroughly checked the numbers, the reporter found that the actual numbers were 5.5% and 26.6% respectively. Stauffer also stated as an example that in the *Hospice Général*, the canton's public welfare services, 14 out of 15 members of one particular department of the institution were French. So by putting forth numbers, Stauffer was actively trying to legitimise a causal relationship between unemployment and economic crisis on the one hand and French borderers on the other.

The MCG is known for its aggressiveness towards the French who work in Geneva, and ever since its formation as a political party it has repeatedly and consistently issued statements linking the economic problems of the canton with purported excessive immigration originating from neighbouring French regions. This is something that anyone in Geneva who follows the news will know, since political slogans such as "ENOUGH with the borderers" ("*Frontaliers ASSEZ*") or "Let us save jobs for the people of Geneva" ("*Réservons les emplois aux Genevois*")[221] have been broadcasted on the internet, in newspapers and on public billboards. Repetition is typical of political communication, and in this case the link between economic problems and

[220] "Eric Stauffer: ses chiffres chocs se révèlent en toc", article available online at: http://www.tdg.ch/geneve/actu/eric-stauffer-chiffres-chocs-revelent-toc-2009-11-02

[221] Taken from the MCG website, available online http://www.mcge.comze.com/fr/charte.php.

French immigration is an idea that the majority of the population in Geneva knows the MCG defends, which would tentatively suggest that whenever something about borderers and economic problems is discussed by a politician, in particular Eric Stauffer, what readily pops to mind is some kind of representation of his opinions about 'French immigrants stealing Swiss jobs'.

Two days after the publication of that article, on November 5th 2009, the same newspaper published another article where the data provided by Eric Stauffer was called, again, into question:[222] the reporter inquired with the department which purportedly employed 14 Frenchmen, and had found out that not a single one of its employees was actually French, which exposed Stauffer's statement as a lie, or at least as an error due to his ignorance of the facts. The article reports that to this, Stauffer first replied by correcting his statement saying that there were actually 13 Frenchmen out of 14 employees in that service, but that, when confronted again with the truth, he replied that the reporter's data was inaccurate and gave the name of the employee of the public welfare service who had given him the data. Upon inquiry, the employee revealed that what had happened is that he had participated to a professional meeting in that service where 14 out of the 15 people that were present were in fact Frenchmen. However, instead of withdrawing his argument – since people who attend a meeting in a given institution are not necessarily employees of that institution – Stauffer persisted: "Once again, you can well witness their stranglehold! This has to do with what people feel! [and not with numbers]" ("*Encore une fois, vous voyez bien leur mainmise! C'est par rapport au ressenti des gens!*"). His whole campaign being directed at portraying French borderers as the people responsible for the Genevan economic crisis, he tried to defend himself by refuting the numbers found by the journalist, and by invoking time constraints in political debate, which, according to him, do not allow him to fully express the details of his arguments.

Let us move to the interesting bit. On the right side of the article explaining how Stauffer manipulated the numbers, the journalist reports four testimonies of MCG supporters, who are obviously sympathetic to Eric Stauffer's cause. Three of them are of particular interest (the fourth is actually critical of Stauffer's methods), and here is what they answer, when asked about the manipulation of numbers:[223]

(78) "What do you think about Stauffer's manipulation of numbers?"

---

[222] Article available online: "Comment Stauffer manipule les chiffres"
http://www.tdg.ch/geneve/actu/stauffer-manipule-chiffres-2009-11-04

[223] The question in (78) is my own formulation of what I imagine the reporter has asked the MCG supporters, but (79), (80) and (81) are the actual answers.

(79) "I consider that the numbers given by other parties are even more wrong that those the MCG uses anyway. The latter is right from A to Z."

(*J'estime que les chiffres livrés par les autres partis sont de toute manière encore plus faux que ceux du MCG. Celui-ci a raison sur toute la ligne*).

(80) "The MCG is stigmatised. The press tries to divert its supporters, but we firstly adhere to ideas. Not to numbers."

(*On stigmatise le Mouvement citoyen genevois. La presse essaie de détourner ses sympathisants, mais nous adhérons premièrement à des idées*).

(81) "It is difficult to judge based on one error. It happens to me too: sometimes I am not correct when I provide numbers. Besides, this is not a criterion."

(*Il est difficile de juger sur une erreur. Il m'arrive aussi de ne pas être dans l'exactitude lorsque je donne des chiffres. Et puis, cela n'est pas un critère*).

None of the three found any problem with the fact that Eric Stauffer had used inaccurate data to sustain his negative opinion on French borderers, which was actually the whole point of the article. This is surprising insofar as a seemingly rational attitude to adopt whenever someone's argument is refuted on evidential grounds would be to cast doubt on the standpoint, as it has been shown not to withstand counter-argumentation in the form of factual data. Despite the evidence, they sound even more convinced by the claims of their leader. This could be interpreted as a patent case of self-deception, and as such one could explain it by resorting to the traditional accounts found in philosophical literature (see e.g. Bach 1981 or Clément 2006 for a more recent discussion).[224]

Beyond self-deception, I want to look more in detail at these examples in order to see if a cognitive pragmatic explanation of these testimonies, focused on information-processing, can be of any use. To this effect, I will assume that the example can be interpreted in terms of a contextual constraint bearing on information accessibility, specifically aimed at boosting the relevance of a context C, such that the implications derived from processing C' are disregarded – C' being understood as the context containing information about Stauffer's error.

As said before, anyone with some interest in Genevan politics knows Stauffer's standpoints on the issue of French working borderers, because messages explicitly linking them and economic crisis have been repeatedly issued in a variety of formats (audio, video, paper, etc.). What is

[224] Since I am mostly concerned with manipulative *communication*, I will not get into the details of self-deception mechanisms here, even if such accounts shed light on cognitive aspects that are also relevant to an account of manipulation.

more, such a connexion is very likely to be salient in the cognitive environments of MCG supporters, as they are by definition likely to share their leader's opinions on the matter. I therefore hypothesise that this association (French borderers / economic crisis and unemployment in Geneva) is in all likeliness very accessible in MCG supporters' cognitive environments, especially when the topic of discussion is explicitly (thus accessibly) linked with this sort of asserted connexion. Furthermore, since these are already held beliefs, whenever they are solicited as relevant assumptions in a communicative exchange they are likely to readily pop into mind together in the form of a 'ready-made' context.[225] Now, assuming the high accessibility of this context C, when a piece of information is supplied, weighed against C and found inconsistent, we might surmise that cognitive dissonance is generated, along the following argument: the speakers of (79), (80) and (81) entertain the belief that the proportion of French borderers across employment sectors in Geneva is too high; when the reporter presents them with evidence that this claim, at least in the way that Stauffer presented it in a recent past, cannot withstand factual evidence, they should cast doubt on it, hence the hypothesis of cognitive dissonance: they believe something that is incompatible with what they are told (and this effect should be robust given that the evidence supplied is factual data). In a state of cognitive dissonance, it is believed that the mind will strive for a quick resolution, and I suggest that this is something that has been caused by a previous contextual selection constraint.

Cognitive dissonance calls for the abandonment of one of the dissonant cognitions. In this case, the ready-made context C is so overwhelmingly consistent with the subject's beliefs that an assumption such as

(82) "Stauffer's numbers are wrong and thus his standpoint should be called into question."

will not achieve relevance, not because it shows difficult to process, but because precisely the contextual effects it could achieve are incompatible with the subject's intimate convictions, beliefs and values. Going into the details of the tentative cognitive effects this assumption could have in context C, this is what it would look like: first, since previously held beliefs are too strong, this assumption would not cancel them; second, the assumption will not modify the accuracy of previously held beliefs due to its incompatibility with the latter; finally, the assumption will be unlikely to be kept, as the cognitive system would have trouble accommodating two incompatible beliefs at the same time. In other words, the assumption will be rejected because it will not achieve sufficient relevance in context C. In order to achieve

[225] There are grounds to suspect that communicative stimuli exploiting stereotypical representations also take advantage of such 'ready-made' contexts that are easily and quickly accessed.

relevance and significant cognitive effects, the assumption would have to be processed against an 'unbiased' context C', or a context in which the strength of the assumptions favourable to Stauffer's opinions is low. It is interesting to note the type of rationalisation that the speakers of (79), (80) and (81) mobilise in order to discard the attack represented in (82) and resolve the cognitive dissonance: in (79) the speaker says that other parties do worse, which is a way of lowering the significance of Stauffer's error and making it acceptable, in (80) and (81) doubt is cast on the very significance of factual evidence supplied in the form of numbers, and in (81) the speaker downplays Stauffer's error by acknowledging his own flaws with respect to figures. These are all cognitions which are destined to fight the relevance of the information contained in the reporter's question. Assuming that these people speak in good faith, they are genuine attempts at overcoming the cognitive dissonance induced by the reporter.

Now, even if this account looks like an illustration of constraints on contextual effects instead of processing effort, I still think that there are grounds to defend that prior to any consideration about contextual effect we must reflect upon the reasons why this context C is processed in the first place. I argue that the relevance of the example in terms of processing effort and accessibility is not to be sought in what MCG supporters make of the attack on Stauffer: instead, I suggest that the reason they end up dismissing the attack as irrelevant is that they have no trouble whatsoever in mobilising C. And C is highly accessible because its constitutive assumptions have been previously repeated over and over again, so that whenever communication relating to the topic of French borderers takes place, it pops in immediately. The validity effect in this case – and probably in many cases of propagandistic discourse – is a precondition for accessible contexts: in the particular example of MCG supporters, I think there are grounds to argue that the validity effect had already done its job and that as a result the context C containing assumptions such as "French borderers steal Swiss jobs" and the like is quickly available should the topic of the conversation be about French borderers and what people have to say about it.

The following causal chain of events sums up my discussion of this example:

- the association French borderers/unemployment and its implications have been repeated over a time period by MCG leaders in a variety of formats.
- repeated exposure to these increases their validity in people's minds through the validity effect.
- in parallel, these representations combine into a readily accessible context C when the topic of verbal stimuli requires it.

- whatever input representation about that matter will then be weighed against this constrained but 'ready-made' context that requires little processing effort to be mobilised.
- in case the stimulus' content is compatible with C, the representations it communicates will be found relevant; otherwise, some sort of cognitive dissonance might be generated, depending on the strength of the dissonant cognition (in the example factual data needs serious rationalisation to reduce the dissonance)

III.4.1.1.1.2. Discussion

In this example, the CSC has started to operate upstream, much prior to the utterance of (79), (80) and (81). I argue that such a constraint may underlie the MCG supporters' surprisingly quick disposal of disconfirming data. This, I contend, can be explained by the prevalence and strength of context C over any other context C' and by the fact that C, following low processing effort requirements, is the first seemingly relevant context that is mobilised when the speakers of (79), (80) and (81) react to the question in (78).

Therefore, the strategy of fostering the salience of assumptions by building on their ease of accessibility has a twofold effect: on the one hand it makes those assumptions enter the context C and achieve sufficient relevance in that context, while on the other hands it somehow ensures that other assumptions (in particular the ones which are unfavourable to Stauffer's positions as in (82)), belonging the theoretical context C', are disposed of altogether. In other terms, fostering accessibility of context C also results in diminishing the chances that context C' will be deemed relevant altogether.

I am aware of the apparent simplicity of the example, since it is about individuals who are already convinced by that which the utterance attacks:[226] it thus comes as no surprise that the contextual effects of (82) for MCG supporters are insignificant, and that they would probably be considerably higher for MCG opponents. Moreover, one could object that accessibility is not the only parameter which explains why people end up convinced, but the account provided so far will be augmented with further parameters in the next sections. The properly manipulative nature of the example could also be called into question, since, after all, we don't know whether these people have really been tricked into believing that French borderers steal Swiss jobs. To this, I would respond that even if it is not manipulative, that is, even if the MCG's public figures

[226] In this sense, it would nearly be like trying to convince someone who firmly believes in the existence of God that God does not exist.

actually believe everything they want to convince people of and did not mean to manipulate people in any way, the reasons the addressees end up convinced by these messages have to do with information-processing constraints such as the CSC. The common feature between this example and a more strikingly manipulative one is that the search for the relevance of the target assumption stops once context C is summoned, thereby preventing the subject from accessing a C' which would increase the likeliness of the subject engaging in critical scrutiny of the claim, and this is what any manipulative attempt tries to achieve. Be that as it may, the point of the example was to emphasise that the contextual constraint can be thought of as a precondition explaining why the MCG supporters dismiss factual data so easily; accordingly, whether the example can be judged manipulative or not does not compromise my main claim regarding the way the CSC is gradually enforced through repetition, so that the ready-made context is rendered easily accessible when necessary: this phenomenon somehow functions upstream in the manipulative process, because it is aimed at fostering the accessibility of certain information for further occasions. As such, it works as a facilitator of manipulation, given that it will result in the individual's processing of information within a limited context.

III.4.1.1.2. Emotions, processing effort and the CSC

III.4.1.1.2.1. Emotional refractory state and contextual constraining

As seen in the previous chapter (see §II.3.2.2.), there are grounds to consider that emotional reactions can "operate a preliminary selection on the objects that will be cognitively processed more in depth later on. […] Accordingly, emotions actively partake in the constitution of beliefs since the latter depend on the nature of the information mobilised in cognitive processes" (Clément 2006: 217, my translation).

To define emotions as processes capable of focusing attention is to allow them to play a role in the selection of information; therefore, it amounts to accepting that emotions might play a role in information-processing of communicative stimuli. I argue that emotional responses triggered by communicative stimuli can prompt for contextual selection of a limited set of contextual assumptions (tentatively those which are consistent with the emotional state the addressee is in after exposure to the emotion); in other words, I argue here that emotions can be instrumental to the CSC. More in detail, I would like to propose that emotional responses might have a direct consequence on the accessibility of certain assumptions and the non-accessibility of others.

This is not a new idea, since it corresponds to an intuitive and quite widespread view on the role of emotions in processes of conviction. Early works on propaganda and marketing techniques indeed acknowledge that one can promote products and ideas by speaking to people's hearts, and many works in social psychology (e.g. Cialdini 2009) report the power of emotions. Yet, to

my knowledge, these accounts seldom go beyond the observation that emotions are instrumental to persuasive endeavours otherwise than by postulating some kind of rational/irrational duality.[227] This endows the account with descriptive value, but not with explanatory power, since it usually boils down to the assumption that emotions are powerful because they have been observed to have such power in persuasive endeavours. I will provide a few insights in what follow to enrich this type of account by looking at the ways in which emotions can influence the way we process information, since the latter has been shown to bear strong influence on belief fixation and behaviour adoption. I will also be led to propose some cognitive reinterpretations of classical fallacies in this process.

In order to start explaining this mechanism, let me recall what Paul Ekman (2003) says about the *refractory state* induced by emotions:

> "Suppose someone is furious about having been insulted in public. During his or her fury it will not be easy to consider whether what was said was actually meant as an insult. Past knowledge about that person and about the nature of insults will be only selectively available; only that part of the knowledge that supports the fury will be remembered, not that which would contradict it. If the insulting person explains or apologizes, the furious person may not immediately incorporate this information (the fact of an apology) in his behavior.
>
> For a while we are in a refractory state, during which time our thinking cannot incorporate information that does not fit, maintain, or justify the emotion we are feeling. This refractory state may be of more benefit than harm if it is brief, lasting for only a second or two. In that short window it focuses our attention on the problem at hand, using the most relevant knowledge that can guide our initial actions, as well as preparations for further actions. Difficulties can arise or inappropriate emotional behavior may occur when the refractory period lasts much longer, for minutes or perhaps even hours. A too-long refractory period biases the way we see the world and ourselves."[228] (Ekman 2003: 39-40)

Ekman here points to the very same mechanism that Clément (2006) highlighted in the above quote: emotions may act as filters that prevent the individual from forming representations about some information by operating a selection in the input to cognitive processes. As I defend

---

[227] See for instance Bernays (1928) for early work on propaganda driven by such a distinction.

[228] In a footnote, Ekman draws a parallel between the refractory state and Fodor's notion of encapsulation: "What I am suggesting here is very similar to psychologist Jerry Fodor's account of how information can become encapsulated, by which he meant that information that might not fit with a way of interpreting the world, information the person has stored and knows, becomes inaccessible for a time." (Ekman 2003: 56)

that manipulative communication relies on the CSC, I take it that emotions may act as triggers for the mobilisation of a partial context, by excluding some contextual assumptions that would be relevant otherwise, should the individual avoid the emotional refractory state

III.4.1.1.2.1.1. The 'Sarkozy case'

As a first example, let me build on Sébastien Bohler's analysis of part of an interview of French President Nicolas Sarkozy that took place in February 2007 on the French TV Channel *M6*.[229] At one point, Bernard de la Villardière, the journalist interviewing the President, asks the latter about a controversial aspect of his political programme, namely his intention of lowering the age of criminal majority.[230] In order to respond to that question, Sarkozy recalls the example of a tragic news story, where a woman, Mama Galledou, suffered extreme burning injuries (about 65% of her body was burnt) after some young delinquents threw Molotov cocktails on a public transportation bus in Marseille in 2006. Here is the argument Sarkozy provides in order to sustain his claim that the age of criminal majority should be lowered:

(83) "[About] Mama Galledou, who was burnt in Marseille on that bus: do you think that for this young woman, whose body was burnt to 65%, the fact that the fact that she has been burnt by underage individuals instead of above age people changes anything?" (my translation)

Sébastien Bohler calls in the notion of the emotional refractory period, according to which when you experience a strong emotion, for some time after it, you are unable to consider the relevance of further information that is incompatible with that emotional state; in example (83), Bohler hypothesises in particular that the addressees may fail to represent other assumptions such as "we cannot formulate a law based on how the victims feel", or "perhaps when you are 16 you do not have the same sense of responsibility than when you are 18" because the emotion is too overwhelming to allow the individual to go against it in an attempt to be critical about the received information.

Taking on this analysis, I would like to expand it a little further. Saying that you fail to take into account additional information amounts to saying that you will fail to process that information, but also that you have secured another subset of information that you will take into account. In

---

[229] "5 ans avec… Nicolas Sarkozy", M6, aired on the 25th of February 2007. The analysis of this example, by Sébastien Bohler, is taken from a TV show that is no longer aired, and that used to analyse the role of the media in our societies ("Arrêt sur images", broadcasted on French TV Channel *La 5*). The recording is available on http://www.dailymotion.com/video/x1d8sx_sarko-asi_fun (last accessed, 15.08.2010).

[230] At the time the minimum age to of conviction is 18 years old, and Sarkozy wants to lower it to 16.

other words, an emotionally-loaded verbal stimulus can perform a selection in the nature of the information that the addressee will take into account to process the stimulus. In example (83), this amounts to saying that the addressee will be mobilising a context made of assumptions which are compatible with the indignation or the anger evoked by what happened to this woman, but also that other contextual assumptions, which are incompatible with the story, will be left out, and these are typically pieces of information that would allow the addressee to question the legitimacy of the example, that of the argument as a whole and as a consequence that of the main claim about criminal majority. And in fact this is what happens: in the video, right after Sarkozy utters (83), the camera shifts to the journalist, and we can see that the latter nods and fails to provide any critical objection such as the ones evoked in what precedes.

III.4.1.1.2.1.2. Discussion

In terms of contextual selection, then, what we see in (83) is a constraint on the selection of the contextual information that will be selected for further processing. The emotion triggers, as it were, a constraint that will prevent the addressee from incorporating assumptions which are problematic with respect to the goal Sarkozy is seeking, namely to convince the audience of the legitimacy of the measures he advocates in terms of criminal majority.[231] If the audience had been led to reflecting about the legitimacy of discussing a law in terms of the victim's suffering, that is, should this assumption have integrated the audience's processing of (83), chances are that its output would have revealed a more critical attitude towards what Sarkozy advocates.

Also, in terms of accessibility, (83) shows that emotion-triggering utterances might contribute to making a compatible context C so automatically salient that alternative contexts C' will end up completely disregarded. As a result, context C will be the relevant context of interpretation and context C' will not, given that it will remain unprocessed. The emotional refractory state thus acts as a constraint on the second extent condition of relevance (see §III.3.2.). This is typical of the mechanisms I am accounting for in this chapter: on the one hand reliance of the emotional refractory state allows strengthening some contexts so that they automatically become accessible, and thus require little processing effort; on the other hand, alternative contexts containing assumptions which are incompatible with the emotional reaction induced by the

[231] One could wonder whether the example is manipulative or not. I believe that it is not possible to determine whether this is the case or not. If Sarkozy knows perfectly well that this type of example will prevent his audience from mobilising incompatible assumptions, chances are that his utterance is manipulative (it would be even if he only knew that his chancing of convincing are higher). But this is only something that we can know by asking him (assuming he answers in good faith). Thus, the account of the example I proposed is only meant to explain how it is that the audience may end up convinced, not to establish if (82) was intended as a manipulative move or not.

utterance will be completely obscured in the way Ekman (2003) suggests. For this reasons, I argue that emotionally-loaded utterances may constrain the effort it takes to process them, and that this in turn has strong implications on whether the addressee will accept the utterance or not. In other words, emotions can be used to make sure that the way you understand an utterance determines your acceptance of its content. Refusing to accept a content P is possible only if the set of (contextual) input assumptions of the cognitive process contains assumptions inconsistent with P; emotions prevent the set of non-P-consistent assumptions from making it into the pool of contextual input.

III.4.1.1.2.2. Argumentative appeals to emotion and contextual constraint

Many fallacies that have been identified in Argumentation Theory research over the years seem to fall within the scope of the emotional triggering / information constraining scheme just evoked. Among them we find for instance the *argumentum ad baculum* (or fallacy of the stick), the *argumentum ad misericordiam* (or appeal to pity), the appeal to flattery, many instances of the *argumentum ad consequentiam* (or appeal to the consequences), the *reduction ad ridiculum* (or appeal to ridicule).

Let us take an example of the appeal to pity:

(84) "Professor, you cannot fail me, please! I have just suffered a terrible loss and I was unable to study properly for this exam."

Pragma-dialectics treats this fallacy as a violation of rule number 1, according to which parties should not prevent each other from putting forward or casting doubts on standpoints. While this account describes what happens with such a fallacious move, and why it is fallacious in normative terms, here I would like to examine why such an argument may achieve its goal, and to this effect I will argue that its effectiveness can be accounted for in terms of the CSC. In (84), the speaker is trying to get her addressee to react empathetically so that the emotional load of her tragic loss becomes prevalent in the assessment of what she is asking from her addressee.

If the addressee ends up falling victim to the fallacy, it means that i) the emotional reaction he got was powerful enough to make him revise his judgement (meaning that the context in which he processed the information was optimally relevant), and ii) other information, compatible with his initial judgement on the student's work, was not deemed relevant enough in that calculation. Again, I surmise that this can be explained in terms of a limited contextual subset of assumptions, this in turn being determined by the accessibility of information. Since the emotional reaction can be taken to restrict the set of information to be taken into account, the limited context C against which the utterance will be processed will be the context in which the

utterance achieves relevance; the counterpart of this process is that other information (tentatively constitutive of the context C') will not be accessed in this calculation. I contend that if the addressee manages to mobilise assumptions about the quality of the student's exam, the outcome would potentially be different, in that the context against which the utterance would be weighed would be more elaborate, meaning that it would contain more information in principle. The addressee would then need to weigh the respective relevance and strength of these assumptions, thereby being led to a more complex processing of (84). I will be led further along (see §III.4.1.2.) to complement this explanation with considerations about strength of assumptions.

The argument *ad baculum*, or fallacy of the stick, consists in issuing a threat in order to convince someone of a given standpoint, such as in (85), an example taken from Copi & Cohen:

(85) "The president continues to have confidence in the Attorney General and I have confidence in the Attorney General and you ought to have confidence in the Attorney General, because we work for the President and because that's the way things are. And if anyone has a different view of that, or any different motive, ambition, or intention, he can tell me about it because we're going to have to discuss your status." (Copi & Cohen 1194: 130)

In this example, the standpoint the speaker is putting forth is that members of the administration should continue to support the Attorney General. He defends the standpoint by threatening the audience: whoever disagrees might suffer retaliation as far as their position in the current administration is concerned.

Those who comply choose to follow the speaker's suggestion, namely to support the Attorney General; they have ultimately privileged that solution over the other less convenient one. An example such as (85) – and in fact such as any other appeal to the stick in which some kind of threat is issued – is rather tricky because it does not straightforwardly imply a complete change of belief: those who comply may choose to comply not because they are convinced of the Attorney General's competence, but simply because they fear retaliation. In other words, their decision in this case results from a deliberate and informed choice, since they know all the relevant information to take their decision. In this sense, there would be no point of putting forth the contextual constraint, because both C and C' are represented, and the incompatibility of the assumptions contained in C' with previous beliefs is not resolved by a change in beliefs; rather, the addressee deliberately adopts a behaviour which is incompatible with his beliefs in order to fulfil a separate goal which he deems more important than a public endorsement of his inner convictions.

Still, we might suspect that sometimes the *ad baculum*, combined with an *ad consequentiam* (namely the appeal to desirable or undesirable consequences – and in this case particularly undesirable consequences) might play a role in mechanisms of conviction, to the extent that their emotional load might occult the relevance of other information. In this case, the account of self-deception (see examples (79), (80) and (81) above, see §III.4.1.1.1.) I proposed above might come handy.

The central idea of this sub-section is that many of the fallacies that have been identified over the years and which rely on the exploitation of emotional reactions may be reinterpreted in cognitive pragmatic terms, and more precisely in terms of the constraints they impose onto information-processing mechanisms. In their quality of attention-grabbers, emotions operate selections of information, thereby making some information salient and other information less salient in – or altogether absent from – the context of interpretation. As a consequence, I take it that argumentative fallacies involving emotional constraints of this kind can also be accounted for as generating constraints of communicative inputs.

III.4.1.1.2.3. Flashbulb contexts: how 9/11 can be exploited by the CSC

The above example of the MCG supporters (see §III.4.1.1.1.) provided, according to my account, evidence that repetition of the same message can set the grounds for the CSC to operate on further occasions. However, to the extent that the speakers of (79), (80) and (81) were already supportive of Stauffer's idea, the example could be deemed manipulative on rather shaky grounds, since if they were manipulated, the attempt had probably taken the form of a long-term repetition strategy: the example I discussed was only a consequence of this lengthy process.

The following example also includes a long-term habituation process, and was referred to in the press as the 'Plame affair'. It relies on the progressive establishment of a particularly emotionally-loaded association of two states of affairs that have in fact no connexion. [232]

After the 9/11 attacks on the World Trade Center, American intelligence services set out to find evidence that Saddam Hussein's regime in Iraq was linked to the terrorist attacks in order to justify their intention of going to war with Iraq. Joseph Wilson, a former US ambassador, was sent to Niger by the vice-president's office to investigate if some reports that had previously been gathered, according to which Iraq had bought uranium from Niger to develop nuclear weapons, had any credibility. Wilson came back with no evidence that this was the case, but this did not stop former president George W. Bush from stating in his 2003 State of the Union

[232] The example comes from Didier Maillat (Maillat 2006).

address[233] that "The British government has learned that Saddam Hussein recently sought significant quantities of uranium from Africa".[234] When he saw that his report had been ignored, Wilson leaked the story to the press, which then resulted in a series of retaliatory measures against him and his wife Valerie Plame (her career as an undercover CIA agent was prematurely put to an end, as her secret identity ended up leaked in the press).

The part of the story which is directly relevant to our purposes here is the fact that the Bush administration was looking for a way to establish a strong link between the 9/11 attacks and Iraq, so that their association would become straightforwardly accepted. Such an association became manifest in official speeches and the overall government communication strategy on the matter, who almost systematically established such a link. By the time the USA went to war in March 2003, 76% of the people interrogated in a CNN-TIME poll thought that Saddam Hussein was involved in the 9/11 attacks.[235]

The 9/11 terrorist attacks on the USA constitute a very painful episode of American history with high emotional load. In this respect, they bear a very special status for many Americans: the memories associated to 9/11, according to cognitive psychologists, are "'flashbulb memories' for their highly emotional, meaningful and subjectively permanent nature" (Pickrell *et al* 2004: 353).[236] In addition to the potential influence of such memories on the success of the misinformation effect (Pickrell *et al* 2004), I contend following Maillat & Oswald (2009) that another property of flashbulb memories can be exploited in manipulative communication. Coming back to the Plame affair, there is room to argue that the Bush administration, by intensive repetition of the link they wished to establish between Iraq and Al-Qaeda, was actually trying to take advantage of the emotionally-loaded context of 9/11 became a flashbulb context, readily available to be mobilised for the interpretation of any statement expressing the need to go to war with Iraq. Whenever such an utterance would be processed, chances are that the first set of relevant assumptions that came to mind, thanks to repeated exposure to the stimulus, was

---

[233] Available online at http://usgovinfo.about.com/library/weekly/aasou2003_text.htm. (last accessed 15.08.2010)

[234] A detailed overview of this story can be found online in an article of *The Washington Post*: http://www.washingtonpost.com/wp-dyn/content/graphic/2005/10/29/GR2005102900242.html. (last accessed 15.08.2010)

[235] See "Selling an Iraq-al Qaeda connexion", available here: http://edition.cnn.com/2003/WORLD/meast/03/11/Iraq.Qaeda.link/. (last accessed 15.08.2010)

[236] In other words, information associated to these memories is particularly salient. Pickrell *et al*. (2004) show in particular how post-event information that is misleading can reduce people's memory accuracy for the original event.

one if which the link mentioned previously was salient. The strategy was therefore aimed at sparing the addressee's processing effort in building the context: the flashbulb context functions as the immediately supplied context C, achieves relevance combined with the target utterance and this in turn prevents the addressee from processing an expanded context C' – since relevance has already been achieved by context C.

Similarly to the MCG supporters' example, the government's strategy behind the Plame affair relies on a condition for the success of a later contextual selection constraint: by repeatedly asserting a particular piece of information (i.e., the link Iraq-Al Qaeda) associated to flashbulb memories, they managed to favour the mobilisation of an artificial flashbulb context that would be stored as a readily available, and what is more relevant, context that would be summoned whenever discussions of related information took place. Again, in terms of processing effort, this turns out to be fairly advantageous for the manipulator: the addressee will probably privilege that context over others, by virtue of the low processing effort it requires to be mobilised.

An example I have already mentioned (see §I.2.3.2.1.)[237] illustrates how manipulation may try to exploit emotionally-loaded contexts. Jo Moore, adviser to the British State Secretary Stephen Byers circulated an internal memo on 9/11 telling its recipients that the 9/11 attacks represented a valuable opportunity to "bury bad news". The idea, obviously, was to issue statements that could be damaging to the administration (in this case, notes about councillor's expenses) precisely at that time so that the press would not pay attention to them, as the whole world's attention was focused on New York and the Pentagon.

In this case, what happens is that the 9/11 context is so overwhelming, so easily accessible, that any information that is not relevant with respect to it runs the risk of being discarded as completely irrelevant. In other words, a hypothetical target utterance such as

(86) "The councillors have been spending too much according to these reports and therefore have to reduce their expenses."

will not achieve relevance in a context submerged by information on the 9/11 attacks: on the one hand it does not bear any obvious contextual effects (since it is unrelated to contextual information already there to be picked up), and on the other processing it in order to achieve relevance would require excessive processing effort.

The most interesting aspect of the mechanism at play in this example is that (86) is not misleading *per se*, for it is a public and truthful statement. However, the context in which it is encouraged to be processed will not allow it to achieve relevance, and this explains that it will

[237] This example is also Didier Maillat's. See Maillat & Oswald (forth.) for its complete exposition.

probably be discarded by news reporters, who are otherwise busy dealing with the 9/11 attacks. What happens here is that the context C is so overwhelming that the target utterance stands no chance of staying in the addressees cognitive environment in the long term: even if it is processed, any assumptions its processing might yield will be discarded as irrelevant in the cognitively inescapable context of 9/11. At the same time, C' – the context in which (86) would achieve relevance, for instance by giving the press an occasion to report on the questionable practices of some members of the country's administration – has little chances of being summoned to the extent that C arguably prevails in such a dramatic political context.

The two examples provided in this sub-section illustrated Clément's claim according to which emotions can operate selections in the set of information the cognitive system will process (cf. Clément 2006: 217). The first example, that of the Plame affair, showed how a ready-made context can be rendered highly accessible by exploiting flashbulb memories associated to 9/11 together with the instilment of a specific belief in the audience (in this case the link between Iraq and Al-Qaeda). Over time, this strategy is meant to constrain the contexts against which related information will be processed in the future. I have argued that this strategy is advantageous in terms of information-processing because it renders said context easily accessible, and therefore increases its chances of being selected as a relevant context of interpretation for utterances about 9/11-related issues. The second example also relies on the emotional properties of the 9/11 context, but this time it illustrates how the overwhelmingly powerful presence of a given context may render other target assumptions irrelevant, thus tentatively leading the addressees to discard them. The overly accessible context is too strong and as such overtakes the context in which the target utterance should actually be processed, so much so that the strategy could also be interpreted as a way of keeping a context unprocessed or irrelevant (see §III.4.2.).[238]

Having concluded this section on constraints on processing effort, I now turn to examine the type of influence that constraints on salience of assumptions can have in contextual selection constraints on information-processing.

III.4.1.2. Fostering strength of assumptions in context C

Under the standard RT view, "factual assumptions can be entertained with greater or lesser confidence; we think of them as more or less likely to be true" (Sperber & Wilson 1995: 75). The strength of an assumption denotes the degree to which we believe that it is likely to be the case. In other words, our cognitive system is equipped to ascribe degrees of likeliness to whatever

[238] More generally, I will be led to argue that these mechanisms nearly always go on a par, since they bear on a twofold process.

propositional content we can represent, and this entails that we hold certain assumptions as being more likely the case than others.

Interestingly, Sperber and Wilson postulate that the degree of strength of a given assumption can be envisaged similarly to its degree of accessibility: the strength of an assumption is taken to be a result of its processing history, just like the degree of accessibility of an assumption is a result of its processing history (i.e., whether it has been repeatedly processed or not). The difference between both lies in that the parameter of strength calls into examination the way in which the assumption has been acquired in order to determine its initial strength. In this respect, assumptions formed on the basis of perceptual evidence are likely to be strong since our perceptual mechanisms have presumably evolved in such a way that they are attuned to providing reliable representations of the world. In communicative scenarios, given our natural tendency to trust people and in the absence of counter-evidence, explicit contents are likely to be ascribed a high degree of strength. However, we could cautiously generalise this as Sperber and Wilson do: "assumptions based on the acceptance of somebody's word have a strength commensurate with one's confidence in the speaker" (Sperber & Wilson 1995: 77). As for assumptions acquired through inferential processes, they will be as strong as the weakest of the premises from which they are derived, but it should also be noted, following Sperber *et al*. (1995), that people usually trust their own inferential processes, so much so that they are "nearly-incorrigible cognitive optimists. They take for granted that their spontaneous cognitive processes are highly reliable and that the output of these processes does not need re-checking" (Sperber *et al*. 1995: 90). Taking this last comment into account, we are led to conclude that people will usually associate a high degree of strength to the output of their cognitive processes.

Since there are different ways of acquiring assumptions, since the specificities of the sources allow for variable degrees of strength in the final output and since variability in strength can result in variability of information salience, there is room to suspect that there are communicative strategies meant to affect the strength of contextual assumptions so that they become salient enough as candidates to achieve contextual relevance. The same way that assumption accessibility can be tweaked to foster salience, as previously seen, so can the strength of assumptions be increased. In what follows I examine different strategies tailored to serve this purpose.

III.4.1.2.1. The argumentum ad verecundiam

The so-called "misplaced appeal to authority" (cf. van Eemeren & Grootendorst 2004: 159), originally called *argumentum ad verecundiam* by John Locke, is a vivid illustration of how the CSC might operate on the strength of contextual assumptions. Additionally, it also constitutes a

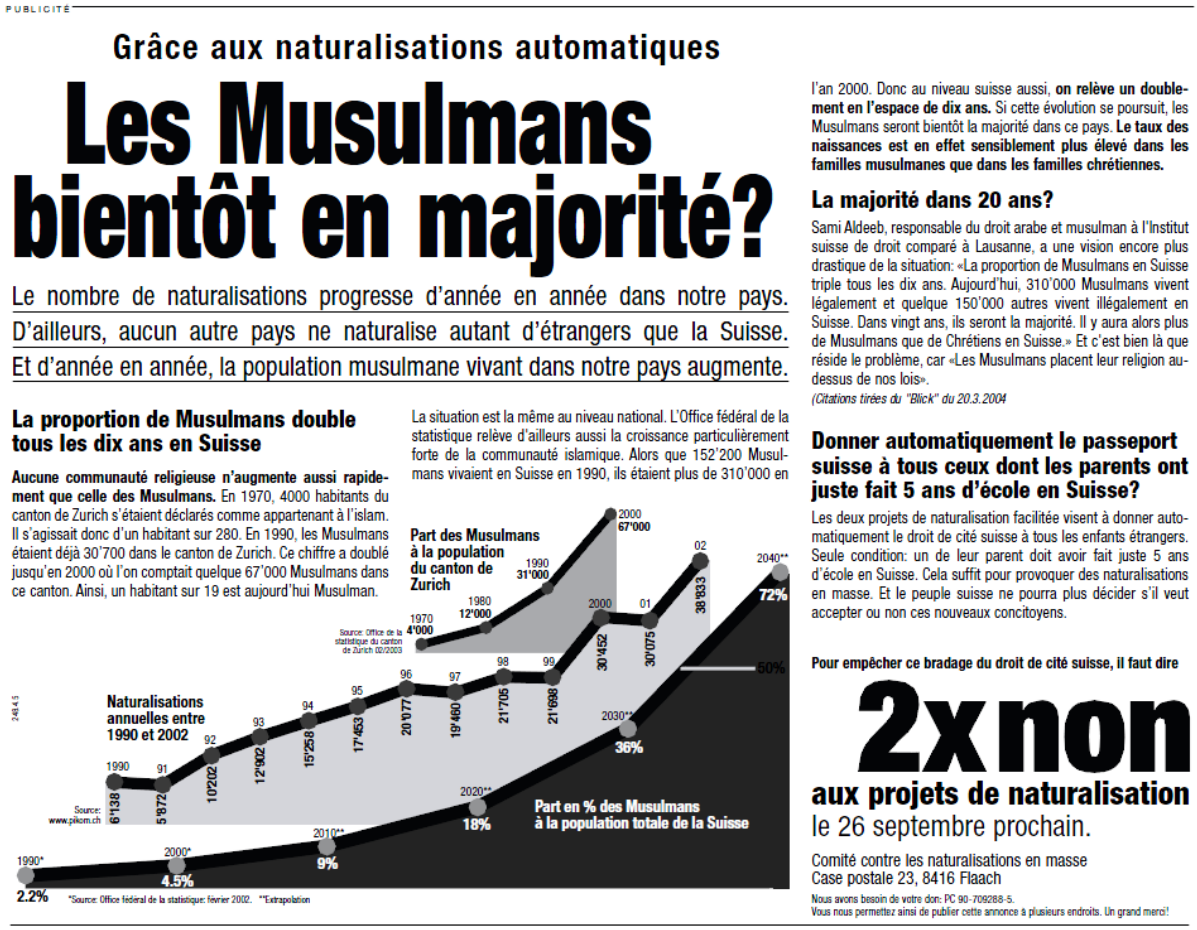
clear illustration of a communicative strategy aimed at tricking an addressee's epistemic vigilance by misleading him into believing that the source of the information he is considering is actually reliable.

As developed by Clément (2006: 75-77), there are reasons to suspect that we are naturally more inclined to trust what others say than to adopt a systematic sceptical position towards communicated information. As a matter of fact, it usually pays off to trust those we think know better than us, in many contexts, both in practical and more abstract matters (it seems perfectly rational to ask a local person for directions in a foreign city and also for children to trust what their school teacher says). The Milgram experiment (see §II.2.1.1.) scientifically attested the strength of this mechanism through a dramatic illustration, and I have pointed out that communicative strategies such as the *ad verecundiam* exploit this natural tendency.

In terms of information-processing, presenting evidence stemming from a reliable authority such as an expert or people we are simply likely to trust is a way of increasing the strength of whatever content such a knowledgeable source might communicate. Nonetheless, granted that trust in authoritative and expert figures is indeed a natural stance, it does not follow that every appeal to authority is legitimate. Argumentation theory, over the years, has been led to distinguish sound from fallacious appeals to authority and has reached consensus on the idea that the difference is contextual, to the extent that in order to legitimately support some standpoint the authority must be an expert in the field to which the standpoint belongs (and should also be agreed upon as an acceptable judge on the matter by the community of arguers). Van Eemeren and Houtlosser show in this respect that for an appeal to authority to be legitimate requires that the acceptability of said authority in the argumentative exchange be acknowledged (either explicitly or implicitly) by the arguers; the consensual nature of the appeal to authority is what makes it valid or acceptable in a given critical discussion (see van Eemeren & Houtlosser 2003). The fallacious argument occurs when such consensus has not been secured, or when the arguer distorts the words of the authority that was agreed upon as a legitimate authority. In these last two cases, and granted that the speaker is aware of the failure to meet these requirements in adducing the testimony of the authority, the fallacy will be manipulative; thus, it can be interpreted within the CSC model, by postulating that there is a C context containing the assumptions relevant to the comprehension procedure and C' context C', which is typically the context containing information allowing for doubt to be cast on the legitimacy of the authority or the faithfulness of the speaker's report of what the authority has said.

III.4.1.2.1.1. The 'UDC anti-naturalisation case'

Let me take an example of an *ad verecundiam* that could be said to further manipulative goals.[239] In 2004, prior to a national vote about the possibility for 2nd and 3rd generation immigrants to benefit from simplified naturalisation procedures, the Swiss far-right political party UDC published a campaign ad in several newspapers entitled "Thanks to automatic naturalisation, the Muslims soon to be in majority?" ("*Grâce aux naturalisations automatiques, les Musulmans bientôt en majorité?*"). Below is a reproduction of the original document:



PUBLICITÉ

Grâce aux naturalisations automatiques

Les Musulmans bientôt en majorité?

Le nombre de naturalisations progresse d'année en année dans notre pays. D'ailleurs, aucun autre pays ne naturalise autant d'étrangers que la Suisse. Et d'année en année, la population musulmane vivant dans notre pays augmente.

La proportion de Musulmans double tous les dix ans en Suisse

Aucune communauté religieuse n'augmente aussi rapidement que celle des Musulmans. En 1970, 4000 habitants du canton de Zurich s'étaient déclarés comme appartenant à l'islam. Il s'agissait donc d'un habitant sur 280. En 1990, les Musulmans étaient déjà 30'700 dans le canton de Zurich. Ce chiffre a doublé jusqu'en 2000 où l'on comptait quelque 67'000 Musulmans dans ce canton. Ainsi, un habitant sur 19 est aujourd'hui Musulman.

La situation est la même au niveau national. L'Office fédéral de la statistique relève d'ailleurs aussi la croissance particulièrement forte de la communauté islamique. Alors que 152'200 Musulmans vivaient en Suisse en 1990, ils étaient plus de 310'000 en l'an 2000. Donc au niveau suisse aussi, **on relève un doublement en l'espace de dix ans.** Si cette évolution se poursuit, les Musulmans seront bientôt la majorité dans ce pays. **Le taux des naissances est en effet sensiblement plus élevé dans les familles musulmanes que dans les familles chrétiennes.**

Part des Musulmans à la population du canton de Zurich
1970 4'000
1980 12'000
1990 31'000
2000 67'000
Source: Office de la statistique du canton de Zurich 02/2003

Naturalisations annuelles entre 1990 et 2002
1990 6'138
91 5'872
92 10'202
93 12'902
94 15'258
95 17'453
96 20'177
97 19'460
98 21'705
99 21'698
2000 30'452
01 30'075
02 38'833
Source: www.pikom.ch

Part en % des Musulmans à la population totale de la Suisse
1990* 2.2%
2000* 4.5%
2010** 9%
2020** 18%
2030** 36%
2040** 72%
60%
*Source: Office fédéral de la statistique: février 2002. **Extrapolation

La majorité dans 20 ans?

Sami Aldeeb, responsable du droit arabe et musulman à l'Institut suisse de droit comparé à Lausanne, a une vision encore plus drastique de la situation: «La proportion de Musulmans en Suisse triple tous les dix ans. Aujourd'hui, 310'000 Musulmans vivent légalement et quelque 150'000 autres vivent illégalement en Suisse. Dans vingt ans, ils seront la majorité. Il y aura alors plus de Musulmans que de Chrétiens en Suisse.» Et c'est bien là que réside le problème, car «Les Musulmans placent leur religion au-dessus de nos lois».

(Citations tirées du "Blick" du 20.3.2004

Donner automatiquement le passeport suisse à tous ceux dont les parents ont juste fait 5 ans d'école en Suisse?

Les deux projets de naturalisation facilitée visent à donner automatiquement le droit de cité suisse à tous les enfants étrangers. Seule condition: un de leur parent doit avoir fait juste 5 ans d'école en Suisse. Cela suffit pour provoquer des naturalisations en masse. Et le peuple suisse ne pourra plus décider s'il veut accepter ou non ces nouveaux concitoyens.

Pour empêcher ce bradage du droit de cité suisse, il faut dire

2x non

aux projets de naturalisation
le 26 septembre prochain.

Comité contre les naturalisations en masse
Case postale 23, 8416 Flaach

Nous avons besoin de votre don: PC 90-709288-5.
Vous nous permettez ainsi de publier cette annonce à plusieurs endroits. Un grand merci!

Figure 5. UDC campaign ad, 2004 (Source: *Le Temps,* 4.09.2004, page 8)

The document contains three graphs with statistics and a text explaining to the reader that simplifying these naturalisation procedures will result in a dramatic increase of Muslims in the country, which is something they implicitly present as an undesirable threat to the Swiss identity but that they merely refer to as "a problem" in explicit terms. Interestingly, the text mentions the testimony of an expert, Sami Aldeeb, who is in charge of the Swiss Institute of Comparative Law at the University of Lausanne. This is what he reportedly says:

[239] The example is taken from Burger *et al*. (2006). The paper contains an interesting Speech act theory-based analysis of a political ad the UDC used in a campaign against simplified naturalisations in 2004.

(87) "The proportion of Muslims in Switzerland triples every ten years. Today, 310'000 Muslims legally reside in the country and some 150'000 others illegally reside in Switzerland. In twenty years, they will be in majority. There will be more Muslims than Christians in Switzerland by then. […] Muslims place their religion above our laws."

("*La proportion de Musulmans en Suisse triple tous les dix ans. Aujourd'hui, 310'000 Musulmans vivent légalement et quelque 150'000 autres vivent illégalement en Suisse. Dans vingt ans, ils seront la majorité. Il y aura alors plus de Musulmans que de Chrétiens en Suisse. […] Les Musulmans placent leur religion au-dessus de nos lois*").

Burger *et al*. note that the name of the expert, Sami Aldeeb, which bears Arabic consonance, licenses inferences that will probably result in a representation of him as someone who does not hold hostile preconceptions about Muslims, since in the absence of more information one could infer that someone with an Arabic-sounding name will be Muslim. This in turn gives more weight to his testimony: since the campaign is directed against Muslims, a supporting testimony emanating from someone who is *a priori* liable not to negatively consider Muslims will confer more credibility to the argument for it will appear as nonbiased. What the authors of the document forget to mention is that Sami Aldeeb is actually a Christian man of Palestinian origin. This first information, if known, would probably modify the inferences we might be able to draw from the mere mention of his name; moreover, they might lead to considering that Aldeeb's testimony could be biased. These are typically pieces of information that would have to be qualified as belonging to the C' context in the manipulative attempt: should they have been manifest, inferences about the speaker's disposition towards the Muslims that were previously licensed from the C context – which was devoid of any such information about Sami Aldeeb's origins and religious persuasion – would have to be modified accordingly, or would perhaps not even have been drawn at all. The fact that Sami Aldeeb is presented through the mention of his professional occupation justifies his expertise in legal matters, and this is hardly questionable. However, the failure to disclose information about his religious persuasion licenses inferences about the evaluation readers might make of Sami Aldeeb's adequacy as an impartial authority: from the consonance of his name readers would probably assume that he is a Muslim, and that, as a Muslim, he is not likely – at least not for dogmatic reasons – to produce negative judgements about that particular religious community. It would thus be much unexpected to hear him formulate such negative judgements, and together with the assumption that he is a credible authority, I suggest that this carries important implications for the readers evaluation of the expert's testimony. I contend in this respect that this therefore fulfils an argumentative function: the combination of the overall message of the ad, which is clearly against the demographic

progression of the Muslim community in Switzerland, with Sami Aldeeb's testimony will trigger an inference of the sort "even a member of that community recognises that they represent a threat to Switzerland", which could be proven wrong should the information about Sami Aldeeb's religious persuasion be made public. This is thus typically an example of information retention destined to fulfil persuasive goals. To the extent that the failure to disclose the critical information is deliberate, the example can also be defined as manipulative.

The mere fact that some information is not disclosed will already potentially trigger inferences which do not correspond to actual states of affairs, but which nonetheless may impact the reader's beliefs about the expert's credibility and the relevance of his testimony. But in addition to the failure to disclose relevant information, the authors of the campaign ad have abused the expert's actual words by reporting them out of the context in which they had been uttered. The authors of the paper (Burger *et al*.) met Sami Aldeeb, who independently confirmed that his words had been taken out of context, and that what he said was actually the result of a reflection about the integration of Muslims, not about their rejection (see Burger *et al*. 2006: 17).[240] In this sense, the campaign ad commits an additional fault because it does not allow the readers to form an adequate picture of the expert's discourse, making the testimony a case of subtle misquoting: even if the words quoted are those that were actually uttered by Aldeeb, there is a mismatch between the context of the original and that of the quote in the ad. This means that the original meaning of the utterance is probably different from the meaning we get from the quote as readers of the campaign ad, since the contexts against which we are to make sense of the utterance are different in both of its occurrences.

The fallacious use of an expert's testimony, in this text, is part of a sub-argument supporting what will be the premise used in a causal argumentation leading to defend the standpoint that simplified naturalisation procedures should not be voted.[241] The argumentative chain is quite difficult to reconstruct in the text because, apart from the standpoint, which is explicitly mentioned at the end of the ad, all other arguments are largely implicit; furthermore, the different sections of the text treat different aspects of the issue which are all instrumental to supporting the standpoint at different levels. Here is however a rough and perhaps overly 'non-

[240] The technique of reporting information in a context which is different from its original context of utterance is a well-known strategy in political discourse, and also a common line of defence when politicians are accused of having said something they do not want to endorse.

[241] This argument could be confused with an *argumentum ad consequentiam*: however, unlike the latter, the example we have in the UDC campaign ad does not defend a truth-conditional standpoint, but a desirable state of affairs (since it urges readers to vote in a certain way). Fallacious exploitations of the causal argument scheme such as the *ad consequentiam* are actually characteristically meant to support the *truth* of a standpoint, which is not the case here.

technical' reconstruction, but for the need of my analysis it should suffice; (88) is the overall causal argument, (89) is the sub-argumentation justifying premise (88b), (90) is the sub-argument in which the *ad verecundiam* is located, and it is meant to justify the factual premise (89b), which is itself instrumental to (88).

(88) a. If we vote for simplified naturalisation procedures, there will be a majority of Muslims in Switzerland.

b. Having a majority of Muslims in Switzerland is undesirable.

c. Therefore, we should not vote simplified naturalisation procedures.

(89) a. Foreigners who place their religion above our laws are undesirable.

b. Muslims are foreigners who place their religion above our laws.

c. Therefore, Muslims are undesirable foreigners.

(90) a. If person P is an acceptable authority in field F, then what P says about F is true.

b. Sami Aldeeb is an acceptable authority in the field of Muslim culture/religion/law.

c. Therefore, what Sami Aldeeb says about Muslim religion, culture and law is true.

d. Sami Aldeeb says that Muslims place their religion above Swiss laws.

e. Therefore, it is true that Muslims place their religion above Swiss laws.

(91) a. People who are not biased against a given culture/religion/law are acceptable experts regarding said culture/religion/law.

b. Sami Aldeeb is not biased against Muslim culture/religion/law.

c. Therefore, Sami Aldeeb is an acceptable expert regarding Muslim culture/religion/law.

(92) a. People who are not inclined to hold negative judgements against the Muslim culture/religion/law are not biased in its favour.

b. Sami Aldeeb is not inclined to hold negative judgements against the Muslim culture/religion/law.

c. Therefore, Sami Aldeeb is not biased in favour of Muslim culture/religion/law.

(93) a. Muslims are not inclined to hold negative judgements against the Muslim culture/religion/law.

b. Sami Aldeeb is a Muslim.

c. Therefore, Sami Aldeeb is not inclined to hold negative judgements against the Muslim culture/religion/law.

(88), (89), (90), (91), (92) and (93) are not formally fallacious in terms of their structure, to the extent that if their premises are true, their conclusions would straightforwardly follow. The problem in this complex argumentation arises from the acceptability of several premises, which ultimately boils down to the acceptability of sub-argument (93b):

- premise (88a) is highly questionable, because the evidence it is based on is itself dubious: the graphs allegedly supporting the claim are mostly projections, and implausible to the extent that if we followed the estimates, we would end up with 144% of Muslims in Switzerland in 2050.[242]
- premise (88b) corresponds to the conclusion (89c); its acceptability thus depends on the acceptability of sub-argument (89), but the acceptability of (89) is itself dubious for two reasons: first, premise (89a) is not self-evident (it would be problematic only if religion was in fact able to override the country's legislation) and second, the acceptability of premise (89b) depends on sub-argument (90), which is itself questionable.
- premise (89b) corresponds to conclusion (90e); its acceptability depends on premise (90b). (90b)'s acceptability depends on the acceptability of (91), whose acceptability depends on the acceptability of (91b), whose acceptability depends on the acceptability of sub-argument (92), and so on and so forth until we evaluate the acceptability of (93b), which is questionable on the grounds exposed above, related to the inference the Arabic-consonance of Aldeeb's name one is licensed to draw regarding his religious persuasion.

III.4.1.2.1.2. Discussion

The *argumentum ad verecundiam*, usually accounted for in mainstream argumentation theories as a misleading appeal to authority determined by the illegitimate status of the expert as an acceptable authority, is here interpreted as a constraint set by the text producer on the strength of certain assumptions that are instrumental to the fulfilment of persuasive – not to mention manipulative – goals. With respect to the example of the UDC 2004 campaign ad, I have tried to show that the problem lies in a particular inference that readers might draw and that is the starting point for a complex argumentative chain based on the acceptability of Sami Aldeeb's authority. Since the critical information – i.e., that Sami Aldeeb is a Christian – is not accessed,

[242] See the lower curve of the graph ("Part en % des Musulmans à la population totale de la Suisse").

the argument from (93) to (88) is licensed. Crucially, assumption (89b) is made extremely salient compared to other assumptions any reader might hold about Muslim demography in Switzerland by virtue of the strength (89b) inherits, in terms of reliability, from its being uttered by an apparently unbiased expert. In this example the information that the expert is Christian is not accessed (therefore belonging to context C'), thereby licensing the inference that Aldeeb is a Muslim, which is not true – but note that the authors of the campaign ad do not bother trying to prevent the misconception, which could be an indication, given the contextual relevance achieved by the information that Aldeeb is a Muslim, of a manipulative attempt. From this follows all the reasoning that ends up rendering assumption (88b) salient to enter context C of interpretation. And in this context, (88b) becomes highly relevant when combined with (88a).

There is no reason to think that the argumentative complexity of my reconstruction of the example is an obstacle for its computation by human minds; probably what I tried to detail here in a complex sequence of arguments and sub-arguments is an automatic and not particularly costly process. However, what I do want to highlight is that the argument would have probably been less effective should the authors of the campaign ad have taken the trouble to mention that the expert voice they were reporting was coming from a Christian. This does not mean that because he is a Christian he would be *ipso facto* negatively prejudiced towards Muslims, but it does mean that his credibility as an expert would differ from the credibility of an expert whom we *a priori* consider not to be biased favourably towards a certain group/culture/religion.

Summing up, the idea behind my exposition of the *argumentum ad verecundiam* is that it operates a constraint on the salience of given information: the assumption supported by the fallacy will be endowed with great strength, for it comes from an expert source; as a consequence, the chances that it will be selected in the context of interpretation are high, this happening to the detriment of other potentially relevant information (typically information that would be contained in the context C'). In the example, this is typically the case with the information about Sami Aldeeb's religious persuasion: the fact that the authors of the ad do not take the trouble to prevent the very plausible inference from the expert's name to his origin and religious persuasion could therefore be taken as an attempt to misleadingly constrain the contextual selection process of their readers so that they end up attributing great credit to the expert.

III.4.1.2.2. The argumentum ad populum

The preceding section advocated that the strength of assumptions can be fostered by presenting the information as emanating from a reliable source, and that this can be unduly exploited by cheating on the reasons for accepting the source as indeed reliable or unbiased. Notably, I have

shown that fooling epistemic vigilance processes was possible through the encouragement of inference derivation by failure to disclose relevant and critical information. Yet, there are two remaining ways in which assumptions may become stronger: first, through perceptual evidence, which I have extrapolated to explicitly communicated contents,[243] and second, through one's own inferential processes.

In this section I will focus on the strength an assumption can inherit from additional information establishing that many people hold the assumption to be reliable. In this sense, it could be seen as the result of a combination of communicated information (such as in "practically everybody adopts a naturalistic stance in psychological research nowadays") and inferential processing on behalf of the individual ("if everybody adopts a naturalistic stance in psychological research nowadays, then such a stance must be correct"). The claim I will be defending is that the *argumentum ad populum* is a way of making information salient, and thus likely to be included in context C of interpretation.

The *argumentum ad populum*, also referred to as 'bandwagon fallacy' or 'appeal to consensus', consists in supporting the truth of a standpoint by resorting to the premise that many people think that said standpoint is the case. The psychological and social reality of this mechanism, as mentioned earlier (see §II.2.1.2.) has been attested in the past in social psychological research on conformity (cf. Asch 1955)[244] and by the principle of 'social proof' (Cialdini 2009); more anecdotally, it is probably fair to say that most of us have witnessed how such a strategy can pay off in terms of the efficiency of reliance on what other people seem to do or think. Advertisers have well understood this natural drive, and that is why many companies advertise their products with slogans highlighting the number of customers who have opted for their products. In the 1970s, McDonald's used to state how many burgers had been sold throughout the world and their signs in front of the restaurants would indicate "Over x billion served", x being updated every 5 billion. Other slogans emphasise that the significant number of people adopting a given behaviour is a strong enough reason to assume that the behaviour is advantageous, as in "Pennsylvania's 5.6 Million Library Cardholders Can't Be Wrong" (Facebook group);[245] should the behaviour not bring any benefit, probably a few of these million people would have noticed

[243] Similarly to Reid's (1785) and Millikan's (1998) view, and granted the assumption that we ordinarily trust people, I will consider that *unless evidence of the contrary*, explicitly communicated information will be held as reliable.

[244] See§II.2.1.2.

[245] Facebook group advocating more funds for Pennsylvanian public libraries. Available on: http://www.facebook.com/group.php?gid=123050057708453 (last accessed 24.08.2010).

and would have cancelled their library cards. Allen Carr, author of a best-selling novel on how to quit smoking, also uses this line of argument when he urges his readers to stick to his method by stating that "it is a complete course in itself and millions of smokers have quit easily by simply reading it" (Carr 2009: xvi). Thus, propositions like 'If millions have managed to quit with Allen Carr's method, why not me?' typically correspond to the type of reasoning that Carr intends to instil in his readers' minds.

Similarly to the *argumentum ad verecundiam*, the *ad populum* increases the strength of the statement it is associated to. Its force stems not from an authority, but from the knowledge that a significant number of individuals take the statement to be correct, desirable and/or advantageous in some way. In other words, all the *ad populum* does is depicting a relevant state of affairs. To some extent,[246] this could be put on a par with the validity effect and the mere exposure effect: even if the statement is not repeated over time, it is still repeated, only that in this case repetition is conceived as a significantly large number of individuals holding the statement to be true. In this respect, it makes little difference whether the statement is repeated over time or repeated individually across a given population. What is more, it is precisely because we have cognitively experienced the effect of (synchronic or diachronic) repetition that the mere verbal instantiation that some statement has been repeated can achieve the intended effect.

Much like the *ad verecundiam*, the *argumentum ad populum* is also a strategy that increases the strength of the target assumption in terms of information-processing. As strength increases, so does salience, and thus so do the chances that the assumption will be selected in context C as relevant input for cognitive processing, at the expense of other less salient assumptions. The mechanisms governing contextual selection, which I exposed above when characterising the CSC are thus also at play with this particular fallacy. By highlighting the consensual or popular nature of a given statement, a speaker is able to boost its salience and consequently its chances of being found contextually relevant. Whatever content embedded within the 'everybody thinks/knows/does P' form associated to the *argumentum ad populum* will tentatively be more salient than others, simply because out of a set of potential contents, that is the one that has been singled out by many people as a valid or advantageous one.

Interestingly, some scholars (e.g. Copi & Cohen 1994, van Eemeren & Grootendrost 2004) link the *argumentum ad populum* to the appeal to emotions:

[246] Following Maillat's (2010) argument.

> "the argument *ad populum*, the appeal to emotion, […] is fallacious because it replaces the laborious task of presenting evidence and rational argument with expressive language and other devices calculated to excite enthusiasm, excitement, anger, or hate." (Copi & Cohen 1994: 128)

The idea is that resorting to consensus is usually accompanied by some kind of either positive or negative emotional emphasis (since the *ad populum* can also be used to reject statements): *ad populum* fallacies are oftentimes used to highlight that a position is overwhelmingly tenable or, on the contrary, untenable, and thus are meant to function as decisive arguments. Their underlying mechanisms probably hinge upon processes of social comparison (the addressee will compare his behaviour/beliefs to that of the majority and notice an inconvenient discrepancy) which may act as motivations for a change in beliefs and/or behaviour. In this sense, the *ad populum* would also be a verbal incentive for such a comparison to be made.

Van Eemeren and Grootendorst are also among those who link the *ad populum* with an appeal to emotions:

> "It is hardly likely that […] the protagonist will state explicitly that he, instead of putting forward any arguments, is only concerned with playing on the emotions of the people who are to be convinced (*argumentum ad populum*)." (Van Eemeren & Grootendorst 2004: 181).

I have discussed earlier (see §III.4.1.1.2.) how emotions, in their quality of attention-grabbers, may be used to focus the addressee's attention on specific information that they render salient, and the relevance of which is likely to be achieved in the context. As a consequence, extent condition of relevance #2, namely the requirement of low processing effort, is also met with such an argumentation insofar as the salience of the target utterance will spare the addressee further allocation of cognitive resources to find a relevant item. The *ad populum*, in this respect, also seems to be a rather clear instance of a CSC based on the exploitation of emotions.

III.4.2. WEAKENING STRATEGIES: LOWERING THE SALIENCE OF CONTEXT C'

Alongside strategies scoping over context C – which do have an indirect effect on context C', to the extent that it is because the context C is retained that context C' is disposed of –, there are diversion strategies which are specifically designed to prevent the addressee from processing context C'. These are deployed to make sure that certain information is either disqualified as a potential candidate for contextual inclusion in context C or left unprocessed altogether. In other words, these are strategies that make sure that the assumptions which belong to context C' stay there.

III.4.2.1. Weakening the salience of C' by making alternative goals salient

Diversion strategies are efficient and commonly used by magicians and illusionists. Following literature on cognitive illusions (see §II.3.1.3.), I contend that manipulators can hope to succeed if they manage to distract their audience's attention from the critical aspects of their message that would lead the audience to spot its fishiness. In particular, this can amount to raise the salience of tasks which are different from the comprehension procedure, so that the addressee does not spend too much time and effort on processing the message so as to understand it. This is a way of lowering the potential salience of the context C'. These strategies operate in communicative exchanges in such a way that they focus the addressee's attention on a specific task, *in addition to the automatic task of making sense of the communicative stimulus*. This is something we have already encountered with the Moses illusion (see §II.3.1.3.1.) and somehow also with example (71) above.

The Moses illusion experiments, as exposed in Erickson & Mattson (1981) consist in making the subjects answer a list of questions. The instructions subjects receive are formulated as follows:

> "The first section of this experiment involves answering questions. The questions will appear one at a time on this screen. After a short time they will disappear. Your job is to read the question out loud, then answer it aloud as fast as you can. With your permission we will tape your answers. Usually a one word answer will be enough. If you don't know the answer or can't recall it right away, just say "don't know." You will occasionally encounter a question which has something wrong with it. For example, you might see the question: "Why was President Gerald Ford forced to resign his office?" The thing that is wrong in this example is that Ford wasn't forced to resign. When you see a question like this, just say "wrong" (or something similar).
>
> So, in summary: you will see a question. Read it out loud, then answer it as quickly as possible. If you don't know the answer, say "don't know." Or if the question is wrong, say "wrong." Following these instructions are three practice questions." (Erickson & Mattson 1981: 542)

If we take a moment to think about the subject's reaction after having read the task, I believe there are good grounds to venture that s/he will be in some kind of 'quiz mode', where her/his knowledge is about to be tested. In other words, I contend that the experimental setting can be perceived by subjects as a kind of playful setting in which they have to take a challenge, to which a whole range of various expectations are associated, such as pressure to perform well, individual eagerness to find the correct solution, emotional excitement, etc. This makes the conditions of information-processing quite different from those of a standard communicative exchange devoid of any such constraints.

In such a context, some sets of assumptions will be deemed highly relevant, but perhaps in a slightly different acceptation of the notion of relevance. Let me introduce here the notion of *relevance to an individual*, which Sperber and Wilson detail in the postface to the second edition of their book. An assumption is said to be relevant to an individual at a given time "if and only if it has some positive cognitive effect in one or more contexts accessible to him at that time" (Sperber & Wilson 1995: 265). A positive cognitive effect is in turn defined as "a cognitive effect that contributes positively to the fulfilment of cognitive functions or goals" (*ibid.*). These two definitions stipulate that some assumptions may be relevant in a given context, not in the sense of contributing to making sense of the stimulus, but in the sense that they further some of the individual's cognitive functions or goals; this is taken to be a general property of our cognitive system, and furthermore a property which is not necessarily dependent upon communicative requirements. Inquiring further about what these cognitive functions or goals may be, we find that these are "other possible contributions to cognitive functioning, involving, for instance, the reorganisation of existing knowledge, or the elaboration of rational desires" (Sperber & Wilson 1995: 266). I argue that the 'elaboration of rational desires', is typically something that may characterise the type of playful setting or 'quiz mode' I believe to be typical of experiments in the Moses illusion vein.

As the expectations linked to the performance of the task rise, providing a relevant answer to the questions can be considered as the fulfilment of a cognitive goal. In this respect, any information contributing to the resolution of the task, that is, any assumption that will be instrumental to the success of the task will have more chances of being selected; additionally, the subject's mind will be set to look for such information in priority. In parallel, those assumptions that are not instrumental to the success of the task will be irrelevant. Under these conditions, therefore, information consistent with the task at hand will be privileged. I contend that this is part of the reason why subjects fail to spot the referential inadequacy explicitly displayed by the substitution of Noah by Moses. They do not spot the anomaly basically because they are busy trying to achieve another, more important, goal.

III.4.2.1.1. The 'Friday the 13th case'

An even more vivid illustration of this phenomenon can be found in an amusing prank performed by French journalist Philippe Vandel in 1999. He took his camera and a microphone and went out on the streets of a French city to ask people the following question:

(94) "Experts have determined that in the year 2000, New Year's Eve falls on Friday the 13th. Does that scare you?"[247]

The question contains an obvious error, since there is no way for a 31st of December to fall on Friday the 13th. And yet, many people fall victim of the trick and answer based on their superstition, failing to notice that the question is flawed. Lexical constraining here lies in the claim that December 31st falls on Friday the 13th. A way of describing the failure to notice the anomaly is to postulate that "Friday the 13th" is not interpreted as any other date, but as a separate lexical item whose main features involve superstitious representations of bad luck. "Friday the 13th" is thus not at all interpreted as "Friday the 21st" would be interpreted, i.e., as a calendar date. The calendar feature is absent of the representation since "Friday the 13th" immediately calls for a set of representations which does not bear calendar relevance. Furthermore, the information is presented as relevant precisely in non-calendar terms, because what the reporter can reasonably be taken to communicate is not 'New Year's Eve falls on Friday the 13th (between the 12th and the 14th)' but something like 'New Year's Eve happens to be an unlucky day this year, according to a widespread superstitious belief'; what would otherwise be the relevance of inquiring about such a special date, were it not for the superstitious connotations of Friday the 13th? I therefore postulate that only representations about bad luck are extracted from the lexical item 'Friday the 13th' in this example. This is compatible with Allott's (2005) account, which makes use of the notion of *ad hoc* concepts (see e.g. Carston 1996, 2002a, 2002b, and also §III.2.2.) to explain how partial representations of lexical items may be induced. It should also be noted that it is the year 2000 we are talking about here, that during the preceding months media have extensively talked about the year 2000 computer bug and the fact that a new millennium was about to begin. This is only speculative, but it could be assumed that, in people's minds, the relevance of being questioned about the matter was quite high at that time.

Moreover, the lexical constraint is backed by an *ad verecundiam* fallacy ("experts have determined that New Year's Eve 2000 falls on Friday the 13th"), which confers scientific grounding to the claim. This, in my view, contributes to keeping the subject's mind away from performing consistency and validity checking. But more interestingly, the subjects are 'in the line of fire', to the extent that they are on camera, with a microphone in front of them and a reporter asking them questions whose answers will be broadcast on national television. Under these potentially stressful conditions, it stands to reason that a significant part of the subject's

[247] The video is available on http://www.dailymotion.com/video/x5k3i_vandel-reveillon-vendredi-13_fun (last accessed 25.08.2010)

cognitive resources will be devoted to providing an accurate answer to the question they are requested to respond to.

III.4.2.1.2. Discussion

Alongside the search for relevance in terms of comprehension, there is an additional task requiring a particular type of answer in order to achieve positive cognitive effects that will be relevant to the fulfilment of a cognitive goal, namely being able to provide an interesting, accurate and relevant answer. I hypothesise that this task competes with the comprehension procedure, so much so that it urges comprehension to be performed quickly and with a minimal expenditure of cognitive resources.

The idea behind this argumentation is quite simple: the reason why manipulation works in these scenarios mostly has to do with the fact that the subject's attention is focused on the resolution of a task that is meant to distract him from being epistemically vigilant. I contend that this type of task is demanding in processing-effort terms, but also in terms of attention and that as a consequence it is likely to mobilise resources at the expense of other cognitive processes. Many mind twisters follow the same principle: they contain an anomalous item and focus the attention of their victim away from spotting the anomaly. Such was the case with the missing franc example (see example (68), §III.0.), and such is the case of many of those, such as the following:

(95) A car is driving with the lights off on a desert road. The lamp posts are all broken, and there is no moon. All of a sudden, a person dressed in black, with a black hood, crosses the street. The car manages to break and avoids the person crossing. How did the driver know that someone was crossing the road?

(96) You have thirty-five cents in two coins. One of them is not a quarter. What are the coins? (taken from Fixx 1977: 78-79)

(97) Some months have 30 days and some months have 31 days. How many months have 28 days? (reproduced from §II.3.3.2., example (67), also taken from Fixx 1977: 80)

In these examples, the relevance we achieve by understanding what is said is quickly reached, not because we do not care about what is said, but because we have more urgent matters to attend to, such as the resolution of the riddle. I suggest that such resolutions become salient cognitive goals that need to be fulfilled in priority and that consequently require allocation of significant cognitive resources, with the result that less resources are available for other cognitive operations such as those performed by our epistemic vigilance systems.

All in all, focusing someone's attention on a specific task, provided s/he is willing to perform it, is a way of keeping context C' away from her/his focus of attention; indirectly, thus, this also

contributes to the addressee's processing the information within the limited context C. What is specific to this manipulative strategy is that the addressee is prompted to try to achieve a more general type of relevance directly related to the task at hand which accordingly makes him achieve 'comprehension-based relevance' very quickly, so quickly in fact that the context C' will presumably not be processed at all.

The mechanism described in this section stipulates that attention can be concentrated on relevant tasks of a perlocutionary nature (such as providing an answer to questions, participating to a scientific experiment, taking a quiz, etc.).[248] This means that, to the extent that they monitor the stimuli we perceive and decide which ones deserve the allocation of cognitive resources, our built-in relevance-seeking devices may be stimulated but also constrained to deal in priority with tasks of different nature, be they communicative or otherwise. If such is the case, this could also happen when the well-being of the organism is in jeopardy, and, by extension, when the communicative exchange makes mention of the 'vital' consequences of what is being discussed for the addressee. As a consequence, the CSC might also be used by manipulators to take advantage of the salience of well-being or life-and-death issues to individuals, and this is also perhaps an indication of the effectiveness of arguments relying on fear, security threats and other life or well-being threatening phenomena. Typically, extremist rhetoric plays with these emotions because they are of great relevance to their audiences. In terms of information-processing, this translates into the idea that the salience of such issues for individuals can prevent addressees from processing information belonging to the context C' – in this case for instance information about the well-foundedness and adequacy to actual states of affairs of the threats.

The attempt to boost the salience of information beyond the comprehension procedure onto alternative tasks whose fulfilment also require cognitive resources is therefore a way of preventing the representation of context C'. This is the reason why I take distracting strategies of this sort to scope over constrains on C'. In terms of the possible argumentative strategies that can fulfil this goal, we find fallacies of the 'red herring' type, i.e., fallacies shifting the focus of attention away from the original argument that was being defended or attacked. I will return to this further down (§III.5.)

[248] And in this sense, those exchanges could be described as requiring perlocutionary cooperation from the addressee (see §I.1.4. on perlocutionary cooperation).

III.4.2.2. The argumentum ad hominem

With an account of the *ad verecundiam* fallacy couched in cognitive pragmatic terms, I have tried to show that the strength of a given assumption could be increased so as to increase its chances of being selected as relevant contextual information to be included in context C. This strategy has a natural counterpart should we consider that strength can, in addition to being boosted, be lowered. And this is precisely the type of effect sought by speakers who resort to the *argumentum ad hominem*.

Argumentation theories usually define the *ad hominem* as a "direct or indirect attack on the opponent" (van Eemeren & Grootendorst 2004: 159-160). More precisely, the fallacy is believed to come in three variants. The first, the *abusive ad hominem*, is an illegitimate direct attack on the target, used to rule out whatever the latter says as invalid, as in example (98):

(98) I wouldn't believe Laszlo if I were you because he is a liar.

The second is the *circumstantial variant*, through which the target's (biased) motivations or self-interest are highlighted to rule out as unacceptable whatever the person utters, as in example (99):

(99) Well, of course the President will vote against this presidential impeachment motion: he *is* the President!

The third variant is the so-called *tu quoque* variant, and is aimed at destroying the credibility of the speaker by emphasising one of her personal inconsistencies, pretty much in an attempt to accuse someone of not practicing what s/he preaches, as in (100):

(100) You're telling me that I'm too rough on my students, but you should hear what your students say about you!

In the standard pragma-dialectical model, the *ad hominem* is usually accounted for as a violation of rule #1, since it restricts the right to advance or to call into question a given standpoint (see van Eemeren *et al*. 2002: 112-113); it proves useful to look more in detail at what the pragma-dialectical school has to say regarding this fallacy:

> "What these three variants have in common is that the discussant who commits one of these variants of the *argumentum ad hominem* fallacy (a) does not tackle the standpoint or the argumentation of the other party, (b) tries to obtain or enforce support for his own position instead, (c) does so by trying to discredit the other party to the discussion, and (d) does so by representing the other party as unworthy of credibility. […] In the *abusive* variant, it is done by calling into question the knowledgeability, intelligence, or integrity of the other party *in general*. In the *circumstantial* variant, an attempt is made to show that personal interests prevent the other party from making an impartial judgment *in the*

> *present case*. In the *tu quoque* variant, an attempt is made to undermine the credibility of the other party by accusing him of lack of consistency if he adopts *this* standpoint in *this* question." (van Eemeren & Grootendorst 2004: 177-178).

What all the variants of the *ad hominem* have in common is the attempt to undermine the target's credibility, so that whatever s/he utters will be rejected as invalid. In information-processing terms, this is equivalent to saying that the *ad hominem* is meant to weaken the strength of a given statement, so that its salience is lowered and its inclusion in context C of interpretation is compromised.

III.4.2.2.1. The 'Trudeau case'

Let me take an example proposed originally by Walton (Walton 1985: 203-204) and first give a pragma-dialectical account of it before complementing the analysis with a cognitive pragmatic take. Walton reports an exchange that took place in the Canadian House of Commons in 1970. The then Prime Minister Trudeau is asked to consider using a government plane to send an information-gathering team for a meeting in Biafra (Nigeria). He responds negatively by saying:

(101) "It would have to refuel in the middle of the Atlantic Ocean…"

Then an MP, Mr. Hees, who is known for his drinking habits, says:

(102) "On a point of order, Mr. Speaker, I bought the plane for the government and I know it can make the flight with the proper stops on the way…"

Mr. Trudeau replies:

(103) "I do not think it would have to stop if the honourable Member went along and breathed into the tank."

Trudeau is asked to justify why he does not want to use a plane, so his standpoint would be something like "I do not consider we should send a team to Biafra with a government plane". The argument in (101) is formulated by Trudeau to introduce some kind of logistic complication, whereby it would be too difficult to manage refuelling over the Atlantic, and this is precisely what Hees calls into question in (102). Before commenting on Trudeau's second intervention, notice that Hees' challenge could itself be considered to be fallacious. He is saying that since he bought the plane, he knows its specifications and, in particular in this context, refuelling issues (which is what Trudeau initially talked about). But in doing that he is asserting his own authority on the matter, and there would be room to question it, since it does not necessarily follow from buying a plane that you are an expert on journey planning and refuelling issues. If his authority does not turn out to be accepted by all parties, then we could evaluate this argument as an instance of ethical fallacy of abuse of authority (which is somehow on a par with the *ad verecundiam* fallacy seen before).

But in (103), Trudeau produces a joke instead of defending and trying to successfully maintain the argument he brought forth, so that his standpoint survives the criticism. In a sense, he is actually opting out of the critical discussion. He does that by shifting the focus of the discussion to the drinking problems of Hees, suggesting humorously that his abusive consumption of alcohol could be used as fuel for the plane. This allows him to disqualify the counter-argument, not on the grounds of its soundness, but because it depicts the antagonist as an unreliable, possibly unreasonable interlocutor (to the extent that there is not much you can expect from drunken people in terms of reasonableness). This is therefore an *ad hominem* fallacy, because it is not instrumental to maintaining the critical ideal of reasonableness. Indeed, it violates rule #1, the Freedom rule, by preventing the antagonist from questioning the standpoint. This example in particular is an *ad hominem* in its *abusive* variant, because it is designed to depict the antagonist as unreliable, and inherits its fallaciousness from its failure to address the objection – let alone to invalidate it. If Trudeau wanted to stick to the goal of reasonableness, he could have questioned the legitimacy of the objection, for instance by casting doubt on Hees' objections on the grounds that buying a plane does not necessarily make you an expert on journey planning.

From a pragma-dialectical perspective, the example could also be taken to be a violation of rule #2, the burden of proof rule stating that a party is obliged to defend a standpoint if asked to. Under these conditions, Trudeau is liable to having committed the fallacy of escaping the burden of proof; indeed, Trudeau does not take on Hees' challenge in this exchange and moves on to another point afterwards. The *ad hominem* has therefore allowed him to disregard Hees' objection.

The standard pragma-dialectical treatment of this example describes how the counterargument in (102) is undermined by asserting the speaker's incompetence on non-trustworthiness grounds in (103), but leaves implicit why credibility and trustworthiness are so important in argument processing. Of course, intuitively these are two parameters that require little justification. Yet, with an information-processing-based account, additional insight can be gained in explaining why it is that this fallacious argument is potentially effective. I believe that at least three reasons can contribute to explaining this.

III.4.2.2.2. Discussion

First, as mentioned above, the *ad hominem* modifies the strength of the assumptions emanating from the source depicted as unreliable: the epistemic value of the counter-argument in (102) is decreased by undermining the personal credibility and reliability of its author. As a consequence, the counter-argument becomes less relevant, as it is not taken up in the conversation. The *ad hominem* is thus interpreted here as a weakening strategy, whereby an

assumption whose source is unreliable will be afforded less credit or even eventually disappear in the interpretative process; it consequently amounts to a form of mistrust-generation strategy meant to lower the salience of certain assumptions. This type of account is licensed if we follow the CSC model I have been advocating so far.

The second argument in favour of an information-processing-based account of the *ad hominem*'s effectiveness tackles the importance of the humoristic tone of Trudeau's reply in (103). His joke about Hees' drinking habits actually allows him to instantiate an additional perlocutionary goal in the exchange; in order to understand the joke and experience its humoristic effect, the audience will need first to allocate cognitive resources to its processing; I contend that similarly to the way cognitive trickeries of the Moses illusion type focus people's attention on the performance of a specific task, humour is an attention-grabber whose processing requisitions cognitive resources – at the expense of other cognitive tasks such as consistency checking, argument soundness evaluation and the likes, which are left with little cognitive resources. In other words, the joke can function as a powerful attention-grabber that will significantly affect the audience's likeliness to allocate cognitive resources to critical tasks, thereby decreasing the individual's levels of epistemic vigilance.

A third argument can support the relevance of a cognitive pragmatic account of the *ad hominem* and has to do with the reaction typically induced by the fallacious move in the target of the attack. When attacked personally, be it on solid or shaky grounds, people have a tendency to defend themselves, even more so when their *positive face*, in Goffman's terms,[249] is threatened in front of an audience; in turn, this can be interpreted the instantiation of a new goal that the target of the fallacy will have to pursue in the communicative exchange, a goal the fulfilment of which will require that cognitive resources be allocated to deal with it. As the victim of the *ad hominem* is struggling to come up with a credible defence, the point that was being discussed prior to the *ad hominem* attack runs the risk of being left unattended and most of the times even forgotten. The goal of the exchange is thus no longer directed towards evaluating the acceptability of a claim that was previously made; what matters from that point on is to re-establish the credibility of the victim of the *ad hominem*.

What these three arguments have in common is that they are all aimed at routing the attention of both addressee and audience onto something else than what was being discussed prior to the launching of the attack – i.e., information likely to be part of a context C' the processing of which

[249] Face is defined as "an image of self delineated in terms of approved social attributes – albeit an image that others may share, as when a person makes a good showing for his profession or religion by making a good showing for himself." (Goffman 2005 [1967]: 5).

would be detrimental to whatever the speaker was advocating earlier, and in the case of the example above, information that would weaken his claim that a plane should not be sent to Biafra. When such a strategy is used in a conversation, it amounts to making information relevant to the detriment of other information, hence the red herring designation, common to many argumentative fallacies operating distractions from the initial course of the communicative exchange.

Below is another example that recently made the news in the Swiss press; I contend that it can be analysed according to the CSC because it also combines a humorous constraint with a distraction, but in this case, I contend that the effect of the fallacy is specifically to prevent the addressees from noticing a formal logical mistake.

III.4.2.2.3. The 'Feurer case'

Earlier in 2010, the some members of the Swiss right-wing political party UDC created a gay section in the party in an attempt to give public importance to the minority within the organisation. In response to that initiative, Gregory Logean, president of the young UDC of the Valais canton made a series of aggressive statements in a local radio show expressing his intimate disapproval regarding the creation of such section, portraying them as a "tumour" that needed to be cured with "chemotherapy".[250] In order to respond to the attack, the leader of the UDC gay section, Beat Feurer, answers back with an attack on Gregory Logean:

(104) "As psychology has shown, people who have repressed homosexual tendencies are prone to adopt homophobic behaviour. Perhaps this is a way of explaining the unacceptable and hurtful comments of Gregory Logean." (my translation)[251]

I propose that the argument might be broken down into premises and conclusion as follows:

(105) a. *Major premise*: "People who have repressed homosexual tendencies are prone to adopt homophobic behaviour".

b. *Minor premise (contextually inferred)*: "Gregory Logean has adopted homophobic behaviour".

c. *(Invited) Conclusion*: "Perhaps Gregory Logean is homosexual".

---

[250] The following article published in the Swiss newspaper *Le Temps* relates the story: "La création d'une section gay trouble l'UDC" (2 June 2010), available at: http://www.letemps.ch/Page/Uuid/ef670462-6dbc-11df-94ea-b1120543b320/La_création_dune_section_gay_trouble_lUDC (last accessed, 27.08.2010).

[251] See the official statement here: http://www.gaysvp.ch/cms/2010/06/stellungnahme-zu-aussagen-der-jsvp-uw/ (last accessed 27.08.2010).

The major premise is stated explicitly; the minor premise is contextually inferred, since it constitutes that to which Beat Feurer is responding. The conclusion we derive is part of the humorous effect sought by Beat Feurer. This conclusion constitutes the response to Gregory Logean's attack, as it is presented as a way of explaining why he produced such violent statements.

The argument is complex because it seems to articulate different argumentative strategies. First of all, the appeal to scientific research in psychology meant to back up the major premise could be seen as an *argumentum ad verecundiam*: we have no mention of which research is referred to and furthermore there has been no prior agreement as to the use of that particular authority, and in this sense it would be legitimate for anyone to question its acceptability. The second argumentative strategy employed in the example lies in the use of an *argumentum ad hominem*: Feurer's answer is actually an accusation of inconsistence, and therefore a *tu quoque* variant of the fallacy because it is grounded on the idea that it would be inconsistent for a homosexual speaker to adopt homophobic behaviour.[252] The effect Feurer is therefore probably trying to achieve is the disqualification of Logean as unreliable by virtue of the self-defeating homophobic nature of his words (that is, if we accept Feurer's argument).

Feurer's counterattack is aimed at lowering the credibility and reliability of Gregory Logean, and this is something contextually quite obvious through the appeal to psychological research; however, let us look more closely at the articulation between premises and conclusion. It seems to me that the structure of the formal fallacy named *affirming the consequent* is mapped onto the argument, which is, according to my reconstruction, of the following form (on the right column I provide a natural language illustration of its fallaciousness):

(106)	All P's are Q's	All cats are mammals
	R is a Q	My dog is a mammal
	Therefore, R is a P	Therefore, my dog is a cat

Of course, the argument produced by Feurer is not a clear-cut case of *affirming the consequent*, if only because i) the conclusion is not qualified as certain, but only as probable ("perhaps"), ii) the standpoint (or conclusion) is not even stated and iii) the major premise itself is not universal, as it expresses a tendency. Nonetheless, beyond formal aspects of the argument, the relevance of stating a major premise such as (105a), in the context of Feurer's words, would not be achieved

[252] The *tu quoque* variant of the *argumentum ad hominem* can receive a cognitive dissonance account if we consider that blaming someone for not practicing what he preaches is an accusation of inconsistence. I find in this parallel an additional motivation to draw links between argumentation theory and cognitive psychology.

otherwise than in combination with contextual assumption (105b), which happens to be extremely salient to the extent that it describes exactly what Feurer is reacting against. In other words, I contend that the derivation of (105c) is somehow inevitable in such a context, for I cannot think of any other way to justify the relevance of mentioning (105a). The problem, of course, is that the way the major premise is stated will not allow us to reach the conclusion *otherwise than by relying on* the fallacy of affirming the consequent.

Two elements are of great argumentative interest in this example: first, the use of the fallacy of affirming the consequent and second, the *argumentum ad hominem* in its *tu quoque* variant. Notice however that both are linked in that we cannot understand the latter's functioning as a way of decreasing the credibility of Logean without reasoning according to the former's formally fallacious scheme.

More importantly, I suggest that the CSC model can straightforwardly account for the example by adding an extra explanatory layer regarding the argument's effectiveness. I contend that the reason the formal fallacy is likely to work – i.e., to go unnoticed – is that the *ad hominem* prompts us to focus on the humorous effects of the *tu quoque* attack to the extent that it ridicules Gregory Logean. More precisely, the cognitive efforts that one has to make to attend to the humour-related task in this example become hierarchically more demanding than the efforts required by consistency-checking operations *because it is the very point of (104) to result in an unflattering portraying of Gregory Logean*. Feurer guides our attention onto the mockery and its legitimation through an appeal to science, and in doing so he is driving our attention away from the set of assumptions belonging to the context C'.

III.4.3. THEORETICAL IMPLICATIONS OF THE CSC MODEL FOR THE STUDY OF MANIPULATION

My exposition of manipulative strategies in the previous section brings about an important distinction as far as the operational definition of manipulation is concerned: while certain manipulative strategies directly affect what the addressee understands the message to be, others operate in such a way that it is not comprehension proper, but failure to process certain assumptions that will favour the success of the attempt.

In the first category, some of the strategies are meant to perturb the derivation of meaning proper, so as to mislead the addressee into forming an inadequate representation of speaker meaning. Such is the case for instance of the Moses illusion, the missing franc example and the example of Friday the 13th. Involved in these examples is the addressee's misrepresentation of the meaning of the verbal stimulus: mainly for reasons relating to the fact that their attention is distracted, they end up understanding something which is different from what is actually said –

or end up understanding the message in a partial way. In this sense, we could say that Informative Cooperation (IC) is the level of interaction that is jeopardised, because in those cases the manipulator intends the addressee to form a representation which would be different should the stimulus be processed more critically: in CSC terms, the addressee processes the utterance in a context C instead of a context C' and ends up with a misleading representation. In these cases, hence, the comprehension procedure has been directly affected by the manipulative strategy, to the extent that we can claim that the speaker has been CC-uncooperative.

In several other cases, however, the comprehension of the message is not the crucial condition for the manipulative attempt to be effective because the latter is mostly about making sure that the addressee does not mobilise specific assumptions that would be detrimental to its success. Such is the case of the manipulative usage of *ad populum*, *ad hominem*, *ad verecundiam* fallacies as well as that of fallacies relying on emotion. Most of them involve no particular problem of comprehension; rather they act as distractions meant to orient the addressee's selection of information. In this last respect, they are similar to the first category of manipulative strategies because they also try to prevent the addressee from processing the context C'. Yet, their scope is different from that of the first category, because in those cases what is understood from the message, in terms of content, is not particularly misleading in itself, in the sense that they are devoid of obvious problems of the Moses/Noah mismatch type. In terms of cooperation, it could be argued that this second category of manipulative strategies – which I will shortly describe in relevance-theoretic terms – is about preventing some perlocutionary effects that a representation of context C' would presumably trigger in the addressee's evaluation of the verbal stimulus.

III.4.3.1. Manipulating contextual relevance and relevance to an individual: implications for IC and PC

The attractiveness of the CSC model as a reliable and adequate theoretical account of manipulative communication lies in its capacity to explain how a single principle may account for how both Informative Cooperation (IC) and Perlocutionary Cooperation (PC) may be exploited to further deceptive goals. Its usefulness is thus to be justified through its ability to account for two types of goals that are usually sought to be attained through communication: getting one's message understood and making one's message acceptable enough so that further potential goals instantiated in the exchange are more likely to end up being fulfilled.

The way I describe and explain the effectiveness of manipulative strategies builds on RT's crucial idea that information can be relevant in two ways in and through a given communicative exchange: it can be relevant for the purposes of comprehension and/or relevant for other goals

conversational participants might have. Sperber and Wilson capture this difference as a difference between *contextual relevance* and *relevance to an individual* (Sperber & Wilson 1995 265-266). Under the contextual relevance condition, information will be deemed relevant to an addressee when it contributes to the formation of a presumably adequate representation of what the speaker meant. Relevance, in this acceptation, can be defined as a property of information that is instrumental to the success of IC, since the latter refers to the communicative process by which contextual assumptions are found relevant, authorising *ipso facto* a proper derivation of speaker meaning.

Additionally, information might be found relevant in a given communicative exchange when it contributes to the satisfaction of so-called cognitive functions or goals, which I believe can be put on a par with the satisfaction of perlocutionary goals such as answering a question correctly, being helpful, contributing to a social game of exchanging jokes, mocking a scape goat, so on and so forth. Being able to achieve the type of relevance involved in this second case is still about finding assumptions relevant and in this sense the processing variables (i.e. processing effort and cognitive effect) relating to extent conditions of relevance #1 and #2 still regulate the whole process. The difference is that this kind of relevance is not geared towards IC, but towards PC.

The interesting conclusion that could be drawn from the strategies I described in section 4 of this chapter refers to the possibility that manipulation could exploit both levels concerned with the relevance of communicative stimuli. I would now like to defend the following hypothesis: the manipulative exploitation of the difference between contextual relevance and relevance to an individual is made possible by the fact that both processes, though related, each require dedicated cognitive resources. In other words, IC and PC may both compete for resources. When a perlocutionary goal is instantiated, and provided the addressee decides to try to fulfil it, significant allocation of resources to achieve that goal is likely to follow, sometimes to the detriment of the requirements of the comprehension procedure meant to ensure successful IC (i.e. that the message has properly been understood). In this case, it does not really matter if the message has been properly understood and evaluated as the mind is already led to deal with other tasks which are relevant to the conversational exchange; the task of understanding will presumably lose thoroughness because the attention of the addressee will focus on a separate search for relevance (related to PC).

As the search for relevance depends on processing variables of cognitive effort and effect, I have postulated that manipulation, which crucially operates a constraint on how addressees establish the relevance of information, can operate on both variables. Thus, effort-oriented strategies will try to boost the salience of certain information (constitutive of context C) so that the addressee can represent them with significantly low processing effort. This was the case with the strategies

evoked in §III.4.1.1. and §III.4.1.2. As for effect-oriented strategies, they are geared towards prompting the addressee to select the relevance of given information in such a way that the latter turns out to be instrumental to the fulfilment of different types of goals. This was typically the case in the Friday the 13th story (cf. example (94)) and of the strategies discussed in §III.4.2.1. and §III.4.2.2.

Now that I have detailed operational aspects of the CSC, the distinction between Informative and Perlocutionary cooperation turns out to constitute a useful *descriptive* tool: the different nature of both types of cooperation is compatible with Sperber and Wilson's distinction between contextual relevance and relevance to an individual, because each of these is instrumental to the success of IC and PC respectively. As a consequence, manipulation operating on the comprehension procedure can be described as Informative non-cooperation, and manipulation operating on the perlocutionary aspects of communication can be interpreted in terms of Perlocutionary non-cooperation, both of which obviously operate covertly.

III.4.3.2. Explanatory and descriptive adequacy of the CSC model

The theoretical construction of the CSC model results from an interface between disciplines which have not traditionally been associated together, namely cognitive and social psychology on the one hand and argumentation theory and discourse analysis on the other. The rationale for the elaboration of the CSC was born out of the observation that many aspects of cognitive processing identified by psychosocial and psychological research (see §II.2. and §II.3.) seemed to interpretable as cognitive counterparts of (fallacious) argumentative and discursive strategies that have long been studied by Argumentation Theory and critical approaches to discourse. I then assumed that this kind of interface should be hosted within an information-processing model of communication since only the latter could accommodate the type of constraints that needed to be accounted for, both descriptively and explanatorily. I postulated accordingly that Relevance Theory was a suitable framework to achieve this interface.

The specificity of the notion of relevance makes it an ideal candidate for the explanation of manipulation, if we consider that the success of manipulation basically depends on the addressee's failure to find certain information relevant – information whose representation would be typically lethal for the manipulative attempt, which I defined as the set of assumptions constitutive of context C'. In principle, then, the account suited to the requirements of my definition of manipulation should explain how information could be made relevant and how other information could be rendered irrelevant. And this is something RT is equipped to do through an account of how processing variables such as processing effort and cognitive effect

determine the output of the cognitive processes underlying the treatment of communicative stimuli.

In addition to that, RT makes no restrictions, in terms of their nature, as to the accountability of constraints that can influence cognitive processing: it inherently postulates internal constraints such as economical resource-bound requirements regulating inferential schemes and informational limits specific to people's beliefs systems but also handles external constraints that come with the communicative stimulus, such as representations about the reliability of the source, perlocutionary requirements dictated by the communicative exchange and so forth. For these reasons, RT is an ideal candidate as a framework suited to integrate insights from theories eliciting external influences on language use as well as theories whose prime concern revolves around the properties of cognitive mechanisms. The possibilities afforded by the interface I am here motivating, in principle, are consequently far-reaching as far as descriptive and explanatory requirements of scientific research are concerned.

From a descriptive point of view, the CSC is a *model*, meaning that it is meant to represent in an accurate way its object of study; I have tried to show that the CSC, should we accept the definitional properties of manipulation I highlighted in Chapter I (see §I.3.2.), that such was the case, as the CSC inherently captures the twofold nature of manipulative communication. Furthermore, I have so far not come across any example of manipulation that could not be described in those terms and in terms of the distinction between CC and PC. This is of course not conclusive evidence of the model's descriptive adequacy, but suggests – if we go by Popperian standards – that the model is descriptively valid as long as no counter-evidence has been presented that would invalidate it.

From an explanatory point of view, the CSC has the advantage models of cognitive processing usually have; as its object is a dynamic phenomenon involving specific procedures, it is meant to address relationships of cause to effect, which makes the model a mechanistic one. That means that the model will provide an account of how informational input is selected, combined to other information, processed inferentially by the cognitive system, thus providing a causal explanation of how specific inputs yield specific outputs. This in turn also entails that the model is able to explain but also crucially predict why such outputs are generated and, crucially, why not alternative ones. Which leads me to the last aspect of the CSC that is relevant to establishing its explanatory adequacy: as it is (partly) based on cognitive science, it naturally affords experimental validation of predictions; experiments in the Moses illusion vein have already tried to identify information-processing conditions for the illusion to work, and I expect the operational aspects of the CSC to be testable as well.

A supplementary benefit of the CSC model is that it allows shedding some light on the thorny philosophical question of the difference between understanding and accepting (or believing). Daniel Gilbert and his colleagues started to test some twenty years ago in a series of experiments the accuracy of the Cartesian idea which dissociates understanding and believing: he showed that individuals whose evaluative mechanisms where perturbed by parasitic stimuli ended up being unable to give up the beliefs that they had acquired just by being exposed to information, that is, just by processing it and understanding it (see Gilbert *et al*. 1990, Gilbert *at al*. 1993). Even if the implications of such a conception show difficult to justify (see Clément 2006: 60-79), they are compatible with the idea that we are generally inclined to trust people – and as a consequence to believe *prima facie* that what they tell us is true. The claim should of course be weakened or qualified less radically to the extent that we are also equipped with epistemic vigilance filters that help us combat our initial trusting dispositions in order to avoid being to gullible.

Still, an information-processing model predicting how information selection works based on salience and relevance criteria is well suited to account for at least one of the possible relationships existing between understanding and believing. Comprehension is a process which takes a limited set of input assumptions to unfold and yield an output; to the extent that the input selection must necessarily be constrained in such a process, there are grounds to hypothesise that this will have an impact on evaluative processes presiding to belief-fixation mechanisms the moment we assume that evaluative processes take as input the output of the comprehension procedure. Let me take the cooking analogy again: depending on whether I use peanut flower, wheat flower, chocolate or caramel as ingredients for baking my cake, the result will be different according to my selection of ingredients. As the output of the baking process will be the object of my taste preferences, the initial selection of ingredients is at least partly responsible for my appreciation of the cake: if I like chocolate and that was one of the initial ingredients, I will think the cake is good, or at least better than if it contained caramel (assuming I do not like caramel). I contend that comprehension and belief mechanisms may be interpreted along the same lines: if I take into account assumptions a, b and c into my comprehension procedure, the output x I get will probably be different from the one I would get should I have selected assumptions d, e and f as input material; since the output of the comprehension procedure is precisely what has to be judged as believable or not, and since this output is dependent on the input selection, we can assume that the evaluation in terms of beliefs will also depend on the information that was selected as the input of the comprehension procedure.

I will not get into the details of the technical implications of this question here; assuming that comprehension and evaluation are two different processes raises many questions like knowing

whether they are parallel or sequential, whether they feed each other mutually, etc. I will limit myself to pointing out that in principle any account that is able to adequately assess how information is selected as input to further cognitive processing will be of great value for research on such cognitive processing, be they interpretative or evaluative processes. And I take it that the CSC model is precisely such a thing. To the extent that it also includes an assessment of epistemic properties of information, it will even more be of help when comes the need to understand why people are likely to believe certain information and disbelieve other information.

III.5. CSC AND ARGUMENTATION

III.5.1. FALLACIES ARE NOT NECESSARILY MANIPULATIVE

In the previous sections I have discussed the manipulative exploitation of argumentative fallacies; yet, some clarifications should be made at this point, since my exposition could leave the impression that I take argumentative fallacies to be *ipso facto* manipulative, and this is something I do not claim.

I defined manipulative communication as communication in which the speaker is covertly pursuing a perlocutionary goal, the addressee's unawareness of which being instrumental to its satisfaction (see §I.3.1.3.). Later I specified this by assuming that it was a twofold process leading the addressee to represent some preferred information (in particular context C) and to fail to represent other information (in particular context C'). Those are conditions specifying what happens in manipulative communication from the operative viewpoint of how it works. The first definition, however, gives us one of manipulation's necessary conditions, namely (intentional) covertness.

But it is not necessary to the success of argumentative fallacies that they be intentionally produced, let alone covertly devised as persuasive strategies, for a number of reasons. First of all, in case of formal fallacies for instance, they may be the fruit of mere logical errors, which is enough to disqualify them as manipulative, as it means that they were generated in good faith. Second, fallacies such as informal fallacies of the *ad verecundiam* type may be completely unintentional: the speaker might simply ignore that the authority she is calling upon is unacceptable in a given exchange, for reasons she also does not know of. Third, our cognitive equipment is fallible and imperfect; in order to cope with time and resource constraints, it has developed adaptive strategies which have later on stabilised in the form of heuristics (see §II.3.1.2. and §III.2.); therefore, relying on authority or on what most people do (respectively the underlying motivations of the *ad verecundiam* and the *ad populum* fallacies) is a perfectly natural

and rational way to try to convince someone of something. Producing such a fallacy could therefore equally count as an attempt to deceive or as a natural argumentative strategy.

As a consequence, at least in what regards their definition, fallacies should not be categorised as necessary instances of manipulative communication – even though they can be in case the speaker knows that the particular fallacy she is using will tentatively allow her to convince the hearer of something or to get him to satisfy one of her perlocutionary goals she does not want him to know about. What makes a fallacy manipulative is therefore not something we can tell from its potential effects; instead, it seems that we are led, once again, to reflect on that ever-problematic notion of intention: only if we are able to ascertain that the fallacy was intentionally produced to fulfil a covert perlocutionary goal of the speaker's will we be able to motivate a judgement about the manipulative nature of the fallacy. There is an important consequence to be drawn here: since what makes an utterance manipulative is not intrinsically related to what it produces in term of cognitive reactions, our treatment of manipulative and non-manipulative fallacies need not be fundamentally different. It is thus perfectly sound to consider that fallacies, whether manipulative or not, prompt for the same type of cognitive processing in their addressees. In fact, I will go even further by assuming that the CSC models the standard way of processing information both in argumentative or non-argumentative contexts.

III.5.2. CSC, SOUND AND FALLACIOUS ARGUMENTS

Traditionally, Argumentation Theory has been concerned with assessing the criteria that allow distinguishing sound or valid arguments from fallacious ones. The problem they face is that norms of argument soundness do not translate into norms of argument effectiveness; just because an argument is sound is no guarantee that it will *ipso facto* be convincing. The main reason for this is that effectiveness is highly contextual: the same argument can be convincing for some people, but not for other people. Moreover, the same argument can fail to convince someone at a given time but succeed at another time. All these difficulties make it very arduous to elaborate a solid account of argument effectiveness.

Argument validity, on the other hand, is a notion that can be addressed in many different ways once we agree on a given norm, which is probably the reason why most of the work within the field of Argumentation Studies has focused on the latter: formal logic usually assesses validity in terms of the relationship of necessity encoded in the premise-conclusion deductive articulation; Pragma-dialectics assumes a model of the critical discussion phrased in terms of critical reasonableness and formulates accordingly ten rules ensuring that the argumentative exchange takes properly place, any violation of the rule amounting to a fallacious move; more contextual accounts such as Perelman's stress the importance of an adequate construal of the audience's

values in order to characterise acceptable argumentation (see §II.1.2.3.1.). The main task of these approaches is to decide on the norm that reflects best our ordinary, spontaneous, argumentative practices, and, in particular psychologically plausible norms of argumentation. This last point is usually the most difficult to address because it requires a plausible model of reasoning that can accommodate the complexity and the sometimes very thin line between acceptable arguments and fallacies; quite unsurprisingly, then, mainstream argumentation theory tends to follow van Eemeren and Grootendorst, who consider that "there is no need to have detailed knowledge of all the cognitive processes that play a role in the interpretation of discourse" (van Eemeren & Grootendorst 2004: 74) if we are addressing the structure of argumentation conceived as a verbal and social activity.

The CSC model provides a way to account for the effectiveness of fallacies, to the extent that we consider that their success requires the type of cognitive processing, on behalf of the addressee, required by successful manipulation. The success of fallacious argumentation, just like the success of manipulative communication, requires that the addressee does not identify the argumentative move as being fallacious, for he would reject it should he become aware of the fallacy. In information-processing terms, this means that there is information that the addressee failed to take into account as he processed the information; similarly to manipulative communication, therefore, we can assume that the processing of fallacious arguments can be analysed in terms of a context C (the set of assumptions compatible with the success of the fallacy) and a context C' (the 'potentially damaging' context likely to contain information about the speaker's intention, the inconsistency of the argument, a discrepancy between its content and the addressee's beliefs, etc.). Fallacies can thus be interpreted as a way of constraining the addressee's processing of information so that he does not perform further checking that would jeopardise the fallacy's effectiveness, which is the reason why I have applied the CSC-model to the analysis of some of them in earlier sections (see §III.4.1.2.1, §III.4.1.2.2, and §III.4.2.2).

Some scholars in Argumentation Theory have pointed out the particular status of fallacies that makes them very close to reasoning heuristics: Sally Jackson notes that "fallacies are not incorrect argument schemes or correct argument schemes applied incorrectly, but *failed diagnostic strategies*" (Jackson 1996: 111, my italics) and surmises that their effectiveness can be explained by resorting to economical cognitive strategies: "fallacies are persuasive because audiences use them as *shortcuts* to avoid careful thinking about issues" (Jackson 1996: 104, my italics). Here we find the idea that the success of fallacious argumentation is basically the result of some kind of error of diagnosis, i.e., an error in the evaluation. Moreover, Jackson contends that the effectiveness of fallacies lies in their ability to prompt for cognitive shortcuts instead of thorough and costly cognitive processing: these are precisely the type of economic

considerations that Relevance Theory – and accordingly the CSC model – builds on. Daniel O'Keefe, one of the scholars belonging to the field of Argumentation Theory who has worked the most on persuasion, considers in reference to the rational status of fallacious argumentation that "all we have are fallible procedures", and thus that "certain reasoning practices, practices that might be condemned as normatively indefensible, are being suggested to be rational practices, at least in some circumstances" (O'Keefe 1996: 70). Such a conception makes a direct link between argumentative literature on fallacies and psychological literature on cognitive biases, heuristics and shortcuts evoked above (see §II.3.2.) and meets the rationale I developed for an interfaced model of manipulative communication: in such a model, and consistently with O'Keefe and Jackson's ideas, discursive fallacies come to be defined as having a cognitive counterpart, in that they are linguistic realisations of constraints on information-processing which exploit cognitive biases and heuristics. In this respect, the CSC model I developed in this chapter is an attempt at showing how the type of assumptions we find in Jackson (1996) and O'Keefe (1996) can be developed into a psychologically plausible account of manipulative and/or fallacious communication.

Drawing on Jackson (1996), O'Keefe (2010) but also Clément (2006), it seems that fallacies can be interpreted as tapping into peripheral processing: this is particularly striking for informal fallacies and red herrings such as the *argumentum ad populum*, the *argumentum ad verecundiam* and the *argumentum ad hominem* among others, who are easily explained by reliance on simple rules of thumb urging us to accept whatever the majority accepts, whatever experts say is the case and to reject whatever unreliable speakers say among other shortcuts leading to automatic evaluation of argumentative input. It could be argued that the story is different for formal fallacies, which should in principle concern central processing since they operate on a specific argumentative articulation between premises and conclusion which needs to be reflectively processed to be evaluated; yet, there are two arguments favouring a unified account for both formal and informal fallacies.

First, the reasons why we may be led to miss that a formal fallacy has occurred can be attributed to our inclination to be easily distracted by red herrings and other perlocutionary tasks instantiated in the communicative exchange (see example (104)), and so an interpretation of the success of the formal fallacy in terms of contextual constraint is still plausible: it would boil down to assuming that formal fallacies work because they go unnoticed due to distracting factors.

Second, and this argument is perhaps more fundamental as it is grounded in recent research in evolutionary anthropology, there are reasons to assume that the "arguments used in reasoning are the output of a mechanism of intuitive inference" (Sperber & Mercier forth.), and this has far-

reaching consequences. Mercier and Sperber distinguish intuitive inferences, which are automatic, unconscious, and typically rely on heuristics, from reflective inferences characterising reasoning, which are conscious and typically prompt us to consciously examine the reasons that make us hold as true a given belief. Traditionally, fast and automatic judgements are believed to be the result of intuitive inference (system 1, or peripheral processing route) and arguments are believed to result from reflective inference (system 2, or central processing route); consequently, the notion of argument is habitually linked to that of reasoning. Mercier and Sperber argue instead that arguments, conceived as representations of relationships between premises and conclusions, are also the result of intuitive inference: in other words, arguments are the output of a mechanism of intuitive inference that delivers intuitions about premise-conclusion relationships. As a consequence, they associate arguments with system 1 since they conceive of them as intuitive beliefs about the acceptability of argumentative chains. Reasoning, the most elaborate inferential process, must therefore be broken into two distinct levels involving different levels of awareness. The first level is that of intuitive inference: this first process yields arguments (or intuitive beliefs about premise-conclusion relationships that need not be consciously represented). Combined with premises, these intuitive beliefs then partake in the reasoning process proper, which eventually yields a reflective belief. The second part of the overall process is accordingly not necessary to argue through communication, for the first intuitive level generally suffices to allow for the performance of argumentative moves; it is however needed to provide evaluative judgements of argument acceptability.

The crucial idea of this schematic representation lies in its linking arguments to the output of intuitive processes: as a consequence, what people intuitively consider to be good arguments is not the result of systematic processing, but rather of intuitions about what acceptable arguments are (which are also probably sustained by the individual's own experience of argumentation throughout his life).

Mercier and Sperber's argumentative theory of reasoning elicits the idea that people's construal of good and acceptable arguments (i.e., both formal and informal arguments) is a result of intuitive processes. Reasoning, then, can be dissociated from arguing, because an individual's intuitive beliefs about acceptable arguments are sufficient for this purpose. Such a stance is in turn ideally suited to accommodate the type of account I am defending: potential constraints on the salience of information advocated by the CSC model are constraints that will prevent the addressee, in cases of fallacious communication, from finding reasons to call into question what is being said. Accordingly, the constraint will attempt to make salient information that is compatible with the addressee's intuitive beliefs about argumentation and keep concealed

incompatible and challenging information. In CSC parlance, this means that the constraints in argumentation will make context C salient while inhibiting context C'. Put yet differently, it also means that fallacies will try to prevent the addressee from reasoning about the verbal stimulus he is processing.

The connexion between the CSC model and the argumentative theory of reasoning is made even more salient if we consider Mercier and Sperber's assumptions about the existence of a specific argumentative module, in line with the massive-modularity hypothesis. This dedicated module is specifically assumed to deal with processing of argumentative contents, both in production and reception: "what the argumentation module does then is to take as input a claim and, possibly, information relevant to its evaluation, and to produce as output reasons to accept or reject that claim" (Mercier & Sperber 2009: 154). Embedded in this definition is the idea that argumentative practices crucially rely on information selection, which the CSC model accounts for.

III.5.3. CONCLUDING REMARKS

To conclude the presentation of the CSC model, I would like to add two remarks, one regarding the scope of phenomena it allows to capture, and a more theoretical one expressing the compatibility of the model with mainstream argumentation theory.

The presentation of the model at the 7th ISSA (International Society for the Study of Argumentation) Conference in Amsterdam in July 2010 (cf. Oswald 2010, Maillat 2010) raised an important question in the audience, regarding the type of fallacies that could be dealt with within the model. As it is based on Relevance Theory, it was assumed that the CSC model only deals with fallacies of irrelevance. The latter are fallacies which present evidence that is actually not relevant to the claim they are trying to defend.[253]

The CSC is based on the technical notion of relevance expressed in terms of information-processing, whereby a given piece of information is relevant if it combines in an optimal way processing effort and cognitive effect. That means that the model is in principle able to deal with any piece of information, be it argumentative or not, fallacious or not; this, I contend, is legitimated by the vocation of the account of context the CSC rests on to characterise human communication on the whole. The manipulative and fallacious exploitation of contextual

[253] Typically, this category includes i*gnoratio elenchi*, the *argumentum ad hominem*, the *argumentum at populum*, fallacies relying on emotions, red herrings, and the formal fallacies of affirming the consequent and denying the antecedent.

selection is a natural consequence of the way our standard information-processing mechanisms function, which is fallible and imperfect. Consequently, the CSC model is in principle equipped to deal with any type of fallacy, since, in the end, the success of fallacious arguments rests on the fact that they have not been identified as such. The CSC shows how considerations of relevance orient information selection, and since information selection is perhaps the most crucial of all parameters that are relevant to explain the success of manipulative and fallacious communication, the CSC is ideally suited to shed light on such processes.

The second remark I want to make here concerns the theoretical implications of the CSC for mainstream Argumentation Theory. As mentioned earlier, the latter is usually concerned with criteria of argument validity meant to account for argument soundness. The CSC model is meant to give insights into the cognitive mechanisms underlying argument effectiveness, which is different. The results of my research are thus not to be taken as challenging or refuting existing accounts of argumentation: rather, it is my impression that they complement what is nowadays done in the field. There are in my view several observations that can support this view.

First of all, the analysis of given argumentative productions is not threatened by the account defended here, because both have different goals. The standard Pragma-dialectical model, for instance, is designed to test whether argumentative exchanges correspond to a certain ideal of critical reasonableness: this means that the pragma-dialectical system is designed to allow the production and the evaluation of sound argumentative exchanges. But trying to rely on cognitive research in order to establish why the arguments contained in the latter are effective is a separate question. Analysis in terms of argumentation stages and rule-compliance is therefore still possible and should be preserved: the description of argumentative exchanges and of the extent to which they can be deemed reasonable is left unaffected by the type of account I propose. Nonetheless, it provides insights into *why* argumentative strategies are effective, so it adds an extra layer of analysis that deepens the account of argumentative practices.

Second, there are similarities between the Pragma-dialectical conception of sound arguments and fallacies and a CSC-inspired conception of both categories. The Pragma-dialectical model considers that

> "it is good to observe that although it is generally acknowledged among argumentation theorists that an adequate theory of fallacies presupposes an adequate theory of sound argumentation, it is by no means generally acknowledged that, in addition, these two theories should be connected in such a way that each fallacy has, as it were, its sound "counterpart." The relationship between the fallacy and its counterpart should in fact be such that the reason for the unsoundness of the fallacy is directly related to the reason for the soundness of its counterpart." (van Eemeren & Houtlosser 2003: 289)

In other words, "fallacies do not distinguish themselves as 'different animals' that have certain distinctive features other than their fallaciousness which are shared by all of them and distinguish all of them from their sound counterparts (like dogs are distinguished from cats)" (Eemeren 2010: 199).

This idea of there being two sides to a same coin, as far as sound arguments and fallacies are concerned, is in fact compatible with the operational conception of contextual selection embedded in the CSC model. Since the mechanisms of information-processing it rests on are believed to characterise any communicative exchange, it follows that the mechanisms governing information-processing are the same for sound and fallacious arguments. The difference between both can be assessed in terms of the type of context expansion they call for: fallacies typically block access to the context C', but sound and strong arguments should be able to accommodate this context C', since they should in principle withstand criticism and counter-arguments in a thorough discussion. The difference lies in that the success of fallacies requires that they achieve relevance 'too early' in order for the hearer not to be motivated to process the stimulus further, thereby hindering the likelihood that context C' will be processed.

Moreover, the CSC model predicts that sound and strong arguments are those which afford (and even sometimes encourage) context expansion so as to take into account counter-arguments, additional evidence, etc. The ones that survive all potential contexts C' will thus be the candidates that are more likely to be awarded the status of sound argument. Again, I see a strong connexion with the standard Pragma-dialectical model here because the latter considers that the ideal of critical reasonableness "encourages the systematic submission of the one's party standpoints to the other party's critical doubts" (van Eemeren & Grootendorst 2004: 16). This convergence licences in my view the interpretation of the critical discussion model as a situation in which context expansion is voluntarily fostered.

Finally, a third argument in favour of the complementarity between the CSC model and existing approaches to argumentation is expressed by the widespread take on the contextuality of argument effectiveness. Most approaches hold, as mentioned before, that argument effectiveness depends on contextual factors; following Mercier (submitted) and the rationale of the CSC model, I contend that these contextual factors incorporate considerations of processing effort and cognitive effect, i.e., considerations about the salience of information. The idea is the following: in order to convince, a speaker should foster the salience, in terms of accessibility and strength, of her arguments. Mercier notes that we find a striking resemblance between this idea and something the proponents of the 'New Rhetoric' mentioned over fifty years ago, and who contended the following:

> "One datum in argumentation consists of the agreements available to the speaker as supports for his argument. But this element is so large and capable of being used in so many different ways, that the manner in which one makes use of it is of paramount importance. Accordingly, before examining the use of this datum in argumentation, it is essential that we say something of *the part played by preliminary selection of the elements that are to serve as the starting point of the argument* and by the adaptation of these elements to its purpose." (Perelman & Olbrechts-Tyteca 1969: 115, my italics)

In this quote, Perelman and Olbrechts-Tyteca make explicit mention of the selection of starting points the speaker needs to make in anticipation of the argumentative exchange. When Perelman and Olbrechts-Tyteca talk about the contextuality of argument effectiveness and the hypothesis that a speaker needs to take into account the values, desires and beliefs of particular audiences, there is room to suspect that what they have in mind can actually be interpreted as a context made of assumptions that will be compatible with and instrumental to the success of their persuasive attempt. This is also compatible with the extended Pragma-dialectical model of strategic manoeuvring, in particular with the three aspects that can be of use to win the argument, namely topical potential (i.e., the choice among possibilities regarding standpoints, argument schemes, etc.), audience perspective (i.e., the attempt to frame one's contribution in such a way that agrees with audience demands) and presentational devices (i.e., stylistic and formal features of your argument that are deemed more effective). All of these suppose that the speaker selects some possibilities over others. Both the New Rhetoric and Pragma-dialectics share the basic assumption that the effectiveness of argumentation depends on the notion of selection: the success of one's argument will accordingly crucially depend on the choices the speaker makes. But note that these refer to the production side of communication; with the CSC model, however, the overall analysis will be amplified because it explicitly takes the hearer's perspective by focusing on the cognitive mechanisms underlying what happens on the other side of the communicative process. In this respect, the CSC model provides elements of psychological plausibility that existing argumentation theories have not yet started to assess, so much so that it is perhaps time to start looking for *cognitive* norms of argumentation.

III.6. CSC AND CDA

Critical Discourse Analysis is driven by an ethical concern meant to serve social purposes such as raising the awareness, through the linguistic analysis of public discourses, of power relations in society. I have reviewed some of the most significant challenges that have been addressed to the CDA programme in the past (see §II.1.1.), and highlighted that many of them originated from a critical assessment of the analytical tools that CDA uses, sometimes in superficial and

questionable ways. Additionally, I was led to subscribe to the position, emphasised by Chilton, that "CDA in its later manifestation has made, so far as one can see, no contribution to our scientific understanding of the language capacity in homo sapiens" (Chilton 2005a: 22) and to Hart's claim that "the linguistic theory that [CDA] has applied and the linguistic objects that it has analysed have been fairly limited in scope" (Hart 2010: 5). As a consequence, I was led to present a cognitive model of manipulative communication meant to enrich the CDA perspective on how language could generate consent.

Crucially, I assume that a cognitively grounded account of manipulation is the sort of theoretical model that would increase the chances of CDA making a scientific contribution of social significance for one main reason: if we are able to understand why it is that manipulative communication is effective, why certain fallacies manage to convince despite their departure from validity standards, and moreover if the models used to describe and explain these phenomena are based on a deeper understanding of how communicative input are processed by our brains, then I take it that we will be one step closer to resistance to ideology, social domination and manipulation, in the CDA sense. More generally, I take knowledge on how these phenomena operate – that is, being able to rely on a scientific model endowed with descriptive and explanatory power – to be crucially valuable information if we are to avoid falling victim to them.

Understanding how communicative stimuli may tap into our cognitive information-processing mechanisms can hence be a sensible way of reaching a deeper degree of critical awareness. This is in line with a concern that could be voiced in reaction to the current situation in contemporary and mainstream CDA, which sometimes forgets to consistently address linguistic phenomena and is led instead to focus on larger social issues which are no longer connected to a proper linguistic analysis of texts. In a way, CDA has evolved in a direction which is at times no longer in tune with the original spirit with which its precursors initiated it: Critical Linguistics started off in the 1970's as an enterprise grounded in linguistics meant to address, at a *micro* level, the links between language and thought, in particular the causal relationships between particular linguistic structures and the expression of ideology (which then has implications on social reality).

Now, given that advances in cognitive science have deepened our knowledge of mental processes, in particular in terms of information-processing, there is in principle room to include a cognitive dimension to Critical Discourse Analysis. Such a movement has already stimulated an important strand of research in CDA over the past two decades, mainly under the impulsion of Cognitive Linguistics and the seminal work of Lakoff & Johnson (1980) on the central role of metaphor in the development of thought, and later by incorporating the input of conceptual

blending theory (e.g. Fauconnier & Turner 2002). Critical metaphor research (e.g. Charteris-Black 2004, Chilton 1994, 1996, 2004, Chilton & Lakoff 1995, Goatly 2007, Musolff 2004 among many others) postulates that the way we are led to conceptualise the world as the result of our communicative practices has potentially great implications on our behaviour, and to the extent that we consider, as most cognitive scientists do nowadays, that human behaviour is conditioned by a representation of the world, insights into how these representations are formed will necessarily increase our knowledge about behavioural practices.

CDA used to overlook the link between linguistic structures and states of affairs in the outside world – in particular social states of affairs – as it tended to take for granted, perhaps too rapidly and without further inquiry into cognitive mechanisms regulating this process, the effect of language on thought. Cognitive Linguistics is one of the strong paradigms which has addressed the essential intermediate layer of representation in the analytical process, which provides "the missing link between linguistic representations in text and ideologies or cognitive models which CDA needs in order to explain how discourse can be constitutive of society" (Hart 2010: 25). Chilton's work on political discourse (e.g. Chilton 1994, 1996, 2004, 2005b) is in this respect a clear example of how the concern for a more thorough CDA can take shape into a cognitively grounded analysis with predictive power.

The CSC model presented here takes an alternative theoretical cognitive pathway as it is couched in Relevance-theoretic terms; its main focus is how input information is selected to feed further cognitive processes – in particular processes leading to conviction, persuasion and manipulation –, and to provide an account which can in principle accommodate different sorts of constraints affecting the communicative process and specifically constraints on processing variables such as effort and effect. Reliance on the notion of constraint, I contend, is moreover one way of including the importance of social phenomena in the overall account, to the extent that socially-imposed and socially-generated constraints directly affect information-processing mechanisms (see §II.2.1.). Yet, it should perhaps be pointed out that, the type of phenomena the CSC is concerned with seem to happen somehow *upstream* in the information-processing chain since the model caters for the selection of information which will be processed further along.

Now, as research grounded in Cognitive Linguistics paradigms is foremost concerned with mechanisms of conceptualisation, which arguably take place once information has been properly selected as input to the cognitive system, it could be imagined that an RT-based account and a Cognitive Linguistic-based account could complement each other along the following lines: the CSC model might provide an explanation of how and why constrained subsets of information end up selected as input to the comprehension procedure, while a Cognitive Linguistics-inspired perspective could take over to account for how that information is conceptually integrated and

put to use in view of further operations, notably in terms of the structure and organisation of conceptual models. Of course, the idea of theoretical complementarity along these lines remains to be developed and requires a thorough evaluation which I cannot provide on the grounds of this short discussion. However, the principle of an interface looks fairly addressable, if we follow recent literature at the Relevance Theory/Cognitive Linguistics interface (e.g. Tendahl & Gibbs 2008, Tendahl 2009).

What this short discussion nonetheless suggests is that at least two cognitive directions of research may contribute to a more powerful CDA. One of them, the Cognitive Linguistic tradition, has already shown how analysis could be enhanced by systematic reliance on cognitive aspects of conceptualisation and communicative practices. The second direction of research, which is currently explored within the framework of the CSC model (see Maillat & Oswald 2009, forth. Maillat 2010, Oswald 2010), tentatively represents an additional way of enriching the CDA paradigm by generating explanations of why manipulation can be successful. The analytical toolkit supplied within this framework inherits its descriptive and explanatory adequacy from experimentally gathered evidence within cognitive psychology and experimental pragmatics.

CONCLUSION

I started off this research trying to assess what exactly it could mean, in pragmatic terms, to be uncooperative in an exchange, and concluded from my discussion that envisaging communication in terms of three levels of cooperation is a convenient way of categorising conversational exchanges so as to achieve a finer-grained *description*. However, I have then tried to show – following Sperber and Wilson's criticism of the Gricean framework – that the arguments for an *explanation* of communicative practices in terms of cooperation lacked cognitive plausibility: qualifying conversational exchanges as cooperative endeavours falls short of explaining how exactly a successful exchange of meaning is secured.

I accordingly shifted the focus to information-processing phenomena for two main reasons, the first being methodological and the second theoretical. First, a focus on cognitive processes allows considering all sorts of constraints the organism is exposed to, whether internal or external, which is a stance affording a principled multidisciplinarity amenable to hosting insights from varied fields of research (reviewed in Chapter II) – each of which highlighting the relevance of different sources and types of input to the processes governing comprehension. The overall account is consequently enriched because it can make use of the findings of other disciplines, but relatively to the same object of investigation. Second, from a theoretical viewpoint, such a cognitive perspective on information-processing ideally provides a framework capable of assessing different constraints imposed on comprehension, which is well-suited to start accounting for manipulative communication as I have defined it, i.e., as a way of tampering with people's mechanisms of information selection. The CSC model in this sense provides a suitable framework for the assessment of the operational definition I gave of manipulative communication (see §I.3.2.).

On the practical side, I have argued that manipulative communication could be described as a set of strategies the twofold goal of which is to prompt for the interpretation of the message within a limited context and to drive the addressee's attention away from the manipulative intention or from any set of information the awareness of which would jeopardise the latter. I argued that this could be both described and explained in terms of constraints on information-processing variables such as cognitive effort and effect, two crucial parameters determining the relevance of information in the relevance-theoretic framework I followed here. Coupled with this operational take on manipulative communication is the idea that manipulators need to make some

information look relevant while managing at the same time to keep critical information irrelevant.

These manipulative strategies operate two simultaneous constraints on information-processing. On the one hand, strengthening strategies are meant to increase the relevance of certain information – usually by increasing its salience, in terms of accessibility or strength – so that the victim of the manipulative attempt concentrates her/his attention on that information. This means in turn that strengthening strategies have weakening implications for other sets of contextual information. On the other hand, weakening strategies are deployed to decrease the tentative relevance of critical information – that particular set of information that would be detrimental to the manipulative attempt altogether, or context C' – so that it is left unprocessed. Since weakening strategies allow to discard sets of contextual information, resources are in turn available to process other information, which can also result in a way of increasing the chances that other information is focused on – hence the idea that weakening strategies may also have strengthening implications for other contextual sets of information.

I have shown that research in cognitive psychology offers scientifically significant and quantitatively important contributions to the study of biases, and that the benefits of an integration of these cognitive pathways into the traditional body of research concerned with argumentative fallacies would be high. Both disciplines address very similar phenomena, from different yet highly compatible angles: cognitive psychology is not crucially concerned with communicative phenomena, while Argumentation Theory is; the former instead provides solid and empirically-grounded accounts of cognitive processes, while the latter typically does not. The CSC model is also designed to bridge the gap between both fields since it is meant to host discursive as well as cognitive parameters into the same framework.

A next step in the programme of this research would be a systematic study of argumentative fallacies from a CSC perspective. To the extent that we accept that fallacies tap into our imperfect and fallible information-processing mechanisms (see Jackson 1996, O'Keefe 1996), there are grounds to envisage a systematic study of fallacious reasoning patterns in terms of their cognitive significance and implications. This would potentially enrich Argumentation Theory by providing empirical evidence for an account of the success of fallacious argumentation. In other words, the type of model I outlined here is not meant to replace current views on argument validity or acceptability, but to assess argument effectiveness, which is a different question. The rejection of fallacious arguments is motivated by different sets of standards (e.g. logical, pragma-dialectical, etc.); and yet, conversational evidence tells us that fallacies are committed, used but also accepted time and again. The reasons of this acceptance are the target of a cognitive

argumentative account. The CSC, in this respect, constitutes an attempt at offering psychological and communicative explanations of argumentative phenomena whose properties are otherwise well defined in terms of validity in other theoretical accounts.

There is consequently a need for empirical work. Cognitive psychological studies have begun to address different parameters influencing belief fixation such as trust and vigilance towards lies (e.g. Harris *et al*., forth., Clément in press, Mascaro & Sperber 2009). Yet, what is still to be done is an empirical study of the effects of specific *communicative* and *linguistic* parameters, and in particular linguistic forms and expressions allowing variations in salience. For instance, Saussure[254] voiced the idea that particular phenomena linked to presupposition seemed to occur in some experimental studies of cognitive biases (e.g. Loftus & Palmer 1974); an experimental design meant to test presuppositional accommodation, for instance, would probably be in line with the type of analysis performed in Chapter III. The CSC constitutes an encouraging framework for this, because it affords the same experimental possibilities as most research frameworks in cognitive science (in particular cognitive psychology in the vein of the Moses illusion and the Wason Selection task paradigms).

The implications of such an experimental programme are tentatively far-reaching if we consider that they may provide relevant results as far as cognitive abilities, such as the theory of mind, are concerned. Notably, it would predict that autistic people with mindreading deficits (see Baron-Cohen *et al*. 1985) would be incapable of manipulating other people through communication, for they are traditionally taken to lack the ability to metarepresent other people's mental states; in this sense, both producing and identifying communicative manipulative attempts would be out of their reach, just as the production or interpretation of ironic utterances or the mastery of the effects of impoliteness for that matter. The study of manipulative communication would thus constitute, next to areas of research such as irony and politeness, a valuable addition to the field of cognitive pragmatics, meant to assess the relationship between language and cognitive processes.

To conclude, it should be highlighted that perhaps the fundamental idea of this thesis is that it is very easy for people to miss something they are not looking for. And in particular with respect to manipulative communication, I have made the case through an account of the CSC model that manipulators have at their disposal (but not necessarily in a conscious and reflective way) strategies meant to increase the chances that their addressees will not look for the manipulative intent. Implicitly privileging a bounded rationality view on our cognitive system (in the vein of

[254] Personal communication.

Allott 2007), I was led to conclude that the success of manipulative communication rests on the exploitation of standard information-processing mechanisms; what manipulators actually do is to try to make sure that their targets end up relying uncritically on their spontaneous information selection procedures.

REFERENCES

Allot, N. 2005. "The role of misused concepts in manufacturing consent: a cognitive account". In L. de Saussure & P. Schulz (eds). *Manipulation and Ideologies in the Twentieth Century: Discourse, Language, Mind*. Amsterdam – Philadelphia: John Benjamins, 147-168.

Allot, N. 2007. *Pragmatics and rationality*. PhD Thesis, available on http://www.phon.ucl.ac.uk/home/nick/content/n_allott_phd_thesis_09_2007.pdf.

Allott, N. & Rubio Fernández, P., 2002. "This paper fills a much-needed gap". *Actes de l'atelier des doctorants en linguistique*, Université Paris 7, 97–102.

Amossy, R. 2008. "Argumentation et Analyse du discours : perspectives théoriques et découpages disciplinaires". *Argumentation et Analyse du Discours* 1. (Online journal accessible under: http://aad.revues.org/index200.html)

Andler, D. 2004. "Logique, raisonnement et psychologie". In Andler, D. (ed.). *Introduction aux sciences cognitives*. Paris: Gallimard: 315-405.

Aristotle. 350 BC. *Prior analytics*. Translation by A. J. Jenkinson. (Online transcription accessible under http://classics.mit.edu/Aristotle/prior.1.i.html).

Asch, S. 1955. "Opinions and social pressure". *Scientific American*, 31-35.

Attardo. S. 1997a "Locutionary and Perlocutionary Cooperation: The Perlocutionary Cooperative Principle". *Journal of Pragmatics* 27: 753-779.

Attardo, S. 1997b. "Competition and Cooperation: Beyond Gricean Pragmatics." *Pragmatics and Cognition* 5(1): 21-50.

Attardo, S. 1999. "The place of cooperation in cognition". European Conference of Cognitive Science (ECCS'99) Siena, Italy. October 27-30. 459-464.

Attardo, S. 2003. "On the Nature of Rationality in (Neo-Gricean) Pragmatics." *International Journal of Pragmatics* 14: 3-20.

Audi, R. 1972. "On the conception and measurement of attitudes in contemporary Anglo-American Psychology". 2: 179-203.

Austin, J. L. 1962. *How to do things with words*. Oxford: Clarendon Press.

Axelrod, R. 2006 [1984]. *The evolution of cooperation*. New-York: Basic Books.

Bach, K. & Harnish, R. 1979. *Linguistic Communication and Speech Acts*. Cambridge, Mass.: MIT Press.

Bach, K. 1981. "An Analysis of Self-Deception". *Philosophy and Phenomenological Research* 41: 351-370.

Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyzanowski, M., McEnery, T., & Wodak, R. 2008. "A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press". *Discourse & Society* 19(3): 273-306.

Bandura, A., Grusec, J.E. & Menlove, F.L. 1967. "Vicarious extinction of avoidance behaviour. *Journal of Personality and Social Psychology* 5: 16-23.

Baron-Cohen, S., Leslie, AM., Frith, U. 1985. "Does the autistic child have a 'theory of mind'?". *Cognition* 21 (1): 37–46.

Barth, E. & Krabbe, E. 1982. *From axiom to dialogue. A philosophical study of logics and argumentation*. Berlin/New York: Walter de Gruyter.

Barton S. & Sanford A. 1993. "A case study of anomaly detection: Shallow semantic processing and cohesion establishment". *Memory and Cognition* 21(4): 477-487.

Barton, S. & Sanford, A. 1993. "A case study of anomaly detection: shallow semantic processing and cohesion establishment". *Memory & Cognition* 21(4): 477-487.

Bavelas, J. B., Black, A., Chovil, N. & Mullett, J. 1990. *Equivocal communication*. Newbury Park, CA: Sage.

Bellenger, L. 1997. *La force de la persuasion: du bon usage des moyens d'influencer et de convaincre*. Parius: Esf Editeur.

Bernays, E. 1928. *Propaganda*. New York: Liverlight.

Billig, M. 1999a. "Whose terms? Whose ordinariness? Rhetoric and ideology in Conversation Analysis". *Discourse & Society* 10 (4): 543–58.

Billig, M. 1999b. "Conversation Analysis and the claims of naivety". *Discourse & Society* 10 (4): 572–6.

Billig, M. 2008b. "Nominalizing and de-nominalizing: a reply". *Discourse & Society* 19(6): 829-841.

Billig. M. 2008a. "The language of critical discourse analysis: the case of nominalization". *Discourse & Society* 19(6): 783-800.

Birnbaum, M. 2004. "Base rates in Bayesian inference". In Pohl, R. (ed). *Cognitive illusions. A handbook on fallacies and biases in thinking, judgement and memory*. Hove & New York: Psychology Press, 43-60.

Blakemore, D. 1992. *Understanding utterances*. Oxford: Blackwell.

Blass, R. 2005. "Manipulation in the speeches and writings of Hitler and the NSDAP from a relevance theoretic point of view". In Saussure, L. de & Schulz, P. (eds), *Manipulation and Ideologies in the Twentieth Century: Discourse, language, mind*. Amsterdam: Benjamins, 169-190.

Bornstein, R. & Craver-Lemley, C. 2004. "Mere exposure effect". In Pohl, R. (ed). *Cognitive illusions. A handbook on fallacies and biases in thinking, judgement and memory*. Hove & New York: Psychology Press, 215-233.

Bradac, J., Friedman, E., & Giles, H. 1986. "A social approach to propositional communication: speakers lie to hearers". In McGregor, G., (ed), *Language for hearers*. Oxford: Pergamon, 127-151.

Bratman, M. 1999 [1987]. *Intention, Plans, and Practical Reason*. Harvard University Press (Reprinted, CSLI Publications).

Breckler, S.J. & Wiggins, E.C. 1989. "On defining attitude and attitude theory: Once more with feeling". In Pratkanis, A.R., Breckler, S.J. & Greenwald, A.G. (eds). *Attitude structure and function*. Hillsdale, NJ: Lawrence Erlbaum, 407-427.

Brennan, A. 2003. "Necessary and Sufficient Conditions". Last accessed 02.09.2008 on http://plato.stanford.edu/entries/necessary-sufficient/.

Breton, P. 2004. *La parole manipulée*. Paris: La Découverte.

Burger, M., Lugrin, G., Micheli, R. & Pahud, S. 2006. "Marques linguistiques et manipulation. Le cas d'une campagne de l'extrême droite suisse". *Mots. Les langages du politique* 81: 9-22.

Byrne, R. & Whiten, A. (eds). 1988. *Machiavellian Intelligence*. Oxford: Clarendon Press.

Cacioppo, J. & Petty, R. 1982. "The need for cognition". *Journal of Personality and Social Psychology* 42: 116-131.

Caldas-Coulthard, C. & Coulthard, M. 1996. *Text and Practices: Readings in Critical Discourse Analysis*. London: Routledge.

Carr, A. 2009. *Allen Carr's easy way to stop smoking*. London: Penguin Books.

Carston, R. 1996. "Enrichment and loosening: Complementary processes in deriving the proposition expressed?". *UCL Working Papers in Linguistics* 8: 205-232.

Carston, R. 2002a. *Thoughts and utterances. The pragmatics of explicit communication*. Oxford: Blackwell.

Carston, R. 2002b. "Metaphor, ad hoc concepts and word meaning – more questions than answers". *UCL Working Papers in Linguistics* 14: 83-105.

Chaiken, S., Liberman, A., & Eagly, A. H. 1989. "Heuristic and systematic information-processing within and beyond the persuasion context". In J. S. Uleman & J. A. Bargh. (Eds), *Unintended thought*. New York: Guilford, 212-252.

Chalmers. A. 1982 [1976]. *Qu'est-ce que la science? Popper, Kuhn, Lakatos, Feyerabend*. Paris: Poche essais.

Chalvin, D. 2003. *Du bon usage de la manipulation. Les ressorts cachés de la communication d'influence*. Issy-les-Moulineaux : ESF Editeur.

Chambers, T. 2005. "Misunderstanding Mar Workplace". *The Korea Times*, October 15, 2005. available at http://www.scribd.com/doc/267098/Misunderstandings (last accessed: 26.07.2010).

Charteris-Black, J. 2004. *Corpus approaches to critical metaphor analysis*. Basingstoke: Plagrave Macmillan.

Chilton, P. 1987. "Co-operation and Non-co-operation: Ethical and Political Aspects of Pragmatics". *Language and Communication* 7 (3): 221-239.

Chilton, P. 1994. "'La plaie qu'il convient de fermer…' Les métaphores du discours raciste". *Journal of Pragmatics* 21(6): 181-195.

Chilton, P. 1996. *Security metaphors: cold war discourse from containment to common house*. New York: Peter Lang.

Chilton, P. 2005a. "Manipulation, memes and metaphors: the case of *Mein Kampf*". In Saussure, L. de & Schulz, P. (eds), *Manipulation and Ideologies in the Twentieth Century: Discourse, language, mind*. Amsterdam: Benjamins, 15-43.

Chilton, P. 2005b. "Missing links in mainstream CDA: Modules, blends and the critical instinct." In Wodak, R. & Chilton, P (eds). *A new agenda in (Critical) Discourse Analysis: Theory, methodology and interdisciplinarity.* Amsterdam: John Benjamins, 19-53.

Chilton, P. 2007. "Critical Discourse Analysis", *The Cambridge Encyclopedia of the Language Sciences*, Cambridge University Press.

Chilton, P. & Lakoff, G. 1995. "Foreign policy by metaphor". In Schäffner, C. & Wenden, A. (eds). *Language and peace*. Aldershot: Ashgate, 37-60.

Cialdini, R. 2009. *Influence. Science and practice*. Boston: Pearson.

Clément, F. 2006. *Les mécanismes de la crédulité*. Genève: Droz.

Clément, F. in press. "To trust or not to trust. Children's social epistemology". *Review of Philosophy and Psychology* 1 (2).

Copi, I. & Cohen, C. 1994. *Introduction to logic* (Ninth Edition). New York: Macmillan.

Cosmides, L. & Tooby, J. 1996. "Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgement under uncertainty". *Cognition* 58: 1-73.

Dascal, M. 2005. "The Balance of Reason". In D. Vanderveken (ed). *Logic, Tought and Action*. Dordrecht: Springer, 27-47. (also available online: http://www.tau.ac.il/humanities/philos/dascal/papers/balancefinalissima.pdf).

Davies, B. 2000. "Grice's cooperative principle: getting the meaning across". In Nelson, D. & P. Foulkes (eds) *Leeds Working Papers in Linguistics* 8: 1-26.

Davies, B. 2007. "Grice's cooperative principle: Meaning and rationality". *Journal of Pragmatics* 29: 2308-2331.

Dijk, T. van. 1993. "Principles of critical discourse analysis". *Discourse and Society* 4: 249-283.

Dijk, T. van. 1995. "Discourse semantics and ideology". *Discourse & Society* 6(2): 243-289.

Dijk, T. van. 1996. "Discourse, power and access". In Caldas-Coulthard, C. & Coulthard, M. (eds), *Text and Practices: Readings in Critical Discourse Analysis*. London: Routledge, 85-104.

Dijk, T. van. 1999. "Critical Discourse Analysis and Conversational Analysis". *Discourse & Society* 10 (4): 459-460.

Dijk, T. van. 2005a. "War rhetoric of a little ally: Political implicatures and Aznar's legitimatization of the war in Iraq". *Journal of Language and Politics* 4 (1): 65-91.

Dijk, T. van. 2005b. "Politics, ideology and discourse". *Elsevier Encyclopedia of Language and Linguistics*.

Dijk, T. van. 2006. "Discourse and manipulation". *Discourse & society* 17(2): 359-383.

Dijk, T. van. 2008a. *Discourse and context. A socio-cognitive approach*. Cambridge: Cambridge University Press.

Dijk, T. van. 2008b. *Society and discourse: how context controls text and talk*. Cambridge: Cambridge University Press.

Dijk, T. van. 2008c. "Critical discourse analysis and nominalization: problem or pseudo-problem?". *Discourse & Society* 19(6): 821-828.

Dunbar, R. I. M. 1998. "Cooperation and competition". In: R.Wilson & F.Keil (eds) *MIT Encyclopedia of the Cognitive Sciences*. Cambridge (MA): MIT Press, 201.

Eagly, A. & Chaiken, L. 1993. *The psychology of attitudes*. Forth Worth, TX: Harcourt Brace Jovanovich.

Eemeren, F. van. (ed.) 2009. *Pondering on problems of argumentation: twenty essays on theoretical issues*. Dordrecht: Springer.

Eemeren, F. van. 2010. *Strategic Maneuvering in Argumentative Discourse. Extending the pragma-dialectical theory of argumentation*. Amsterdam: Benjamins.

Eemeren, F. van and Grootendorst, R. 1984. *Speech Acts in Argumentative Discussions. A Theoretical Model for the Analysis of Discussions Directed towards Solving Conflicts of Opinion*. Berlin/Dordrecht: De Gruyter/Foris Publications.

Eemeren, F. van and Grootendorst, R. 1992. *Argumentation, Communication, and Fallacies*. Hillsdale, NJ: Lawrence Erlbaum.

Eemeren, F. van and Grootendorst, R. 2004. *A Systematic Theory of Argumentation. The Pragma-dialectical Approach*. Cambridge: Cambridge University Press.

Eemeren, F. van & Houtlosser, P. 2000. "Rhetorical analysis within a Pragma-dialectical framework". *Argumentation* 14: 293-305.

Eemeren, F. van, & Houtlosser, P. 2002. "Strategic maneuvering: Maintaining a delicate balance". In F. van Eemeren & P. Houtlosser (eds.), *Dialectic and Rhetoric: The Warp and Woof of Argumentation Analysis*. Dordrecht: Kluwer Academic Publishers, 131-159.

Eemeren, F. van & Houtlosser, P. 2003. "Fallacies as derailments of strategic maneuvering: The argumentum *ad verecundiam*, a case in point". in: F.H. van Eemeren, J.A. Blair, C.A. Willard & A.F. Snoeck Henkemans (Eds). *Proceedings of the 5th Conference of the International Society for the Study of Argumentation* (pp. 289-292). Amsterdam: Sic Sat.

Eemeren, F. van & Houtlosser, P. 2007a. "The contextuality of fallacies". *Informal Logic* 27 (1): 59-67.

Eemeren, F. van & Houtlosser, P. 2007b. "Reconnecting Dialectic and Rhetoric: Fallacies as Derailments of Strategic Manoeuvring in Argumentative Discourse". *Anthropology & Philosophy* 8(1-2): 49-67.

Eemeren, F. van, Garssen, B. & Meuffels, B. 2009. *Fallacies and judgments of reasonableness*. Dordrecht: Springer.

Eemeren, F. van, Grootendorst, R. & Snoeck Henkemans, F. 2002. *Argumentation. Analysis, evaluation, presentation*. Mahwah, NJ: Lawrence Erlbaum.

Eemeren, F. van, Grootendorst, R., Snoeck Henkemans, F., Blair, A., Johnson, R., Krabbe, E., Plantin, C., Walton, D., Willard, C., Woods, J. & Zarefsky, D. 1996. *Fundamentals of argumentation theory*. Mahwah, NJ: Erlbaum.

Ekman, P. 1985. *Telling lies: clues to deceit in the marketplace, politics, and marriage*. New York: Norton.

Ekman, P. 2003. *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. New York: Times Books.

Erickson, T. A. & Mattson, M. E. 1981. "From words to meaning: a semantic illusion". *Journal of Verbal Learning and Verbal Behaviour* 20: 540-551.

Evans, J. St. B. T. & Lynch, J. S. 1973. "Matching bias in the selection task". *British Journal of Psychology* 64: 391-397.

Evans, J.St.B.T. 2004. "Biases in deductive reasoning". In Pohl, R. (ed) *Cognitive illusions. A handbook on fallacies and biases in thinking, judgement and memory*. Hove & New York: Psychology Press, 127-144.

Evans, J.St.B.T., Newstead, S.E., & Byrne, R.M.J. 1993. *Human reasoning: the psychology of deduction*. Hillsdale, NJ: Lawrence Erlbaum.

Fairclough, N. & Wodak, R. 1997. "Critical Discourse Analysis". In T. A. van Dijk (ed.) *Introduction to Discourse Studies*. London: Sage, 258-284.

Fairclough, N. 1992. *Discourse and Social Change*. Cambridge: Polity Press.

Fairclough, N. 1995. *Media Discourse*. London: Edward Arnold.

Fairclough, N. 1996. "A reply to Henry Widdowson's 'Discourse analysis: a critical view'". *Language and Literature* 5(1): 49-56

Fairclough, N. 2001. *Language and Power* (2nd edition). London: Longman.

Fairclough, N. 2008. "The language of critical discourse analysis: reply to Michael Billig". *Discourse & Society* 19(6): 811-819.

Fauconnier, G. & Turner, M. *The way we think: Conceptual blending and the mind's hidden complexities*. New York: Basic Books.

Festinger, L. 1985 [1957]. *A theory of cognitive dissonance*. Stanford, California: Standford University Press.

Firth, J. R. 1957. *Papers in Linguistics 1934-1951*. London: Oxford University Press.

Fisk, J. 2004. "Conjunction fallacy". In Pohl, R. (ed). *Cognitive illusions. A handbook on fallacies and biases in thinking, judgement and memory*. Hove & New York: Psychology Press, 23-42.

Fixx, J. 1977. *More games for the super-intelligent*. New York: Popular Library.

Fowler, R. 1991. *Language in the news*. London: Routledge.

Galasinski, D. 2000. *The language of deception. A discourse analytical study*. London: Sage.

Gigerenzer, G. 1991. "How to make cognitive illusions disappear: beyond 'heuristics and biases'". In Stroebe, W. & Hewstone, M. (eds). *European review of social psychology* 2, Chichester: Wiley, 83-115.

Gigerenzer, G. 1996. "On narrow norms and vague heuristics: a rebuttal to Kahneman and Tversky. *Psychological Review* 103: 892-596.

Gilbert, D. T., Krull, D. S. & Malone, P. S. 1990. "Unbelieving the unbelievable: Some problems in the rejection of false information". *Journal of Personality and Social Psychology* 59: 601-613.

Gilbert, D. T., Tafarodi, R. W. & Malone, P. S. 1993. "You can't not believe everything you read". *Journal of Personality and Social Psychology* 65: 221-233.

Giora, R. 2003. *On our mind: Salience, context and figurative language*. New York: Oxford University Press.

Goatly, A. 2007. *Washing the brain: Metaphor and hidden ideology*. Amsterdam: John Benjamins.

Godefroid, J. 2008. *Psychologie. Science humaine et science cognitive (2e édition)*. Bruxelles: De Boeck.

Goffman, E. 2005 [1967]. *Interaction ritual. Essays in face-to-face behaviour*. New Brunswick, NJ: Transaction Publishers.

Greco, S. 2003. "When presupposing becomes dangerous: how the procedure of presuppositional accommodation can be exploited in manipulative discourse". *Studies in Communication Science* 3(2): 217-234.

Grice, H.P. 1957. "Meaning". *The Philosophical Review* 66: 377-88. (Reprinted in Grice 1989).

Grice, H.P. 1975. "Logic and conversation". In P. Cole & J. L. Morgan (eds) *Syntax and Semantics* 3 (*Speech Acts*). New York: Academic Press: 41–58.

Grice, H.P. 1989. *Studies in the way of words*. Cambridge, MA: Harvard University Press

Gu, Y. 2003. "Towards a multiple-goal neo-Grice". *International Journal of Pragmatics* 14 : 45-70.

Gu. Y. 1993. "Pragmatics and rhetoric: a collaborative approach to conversation". In Parret, H. (ed) *Pretending to communicate*. Berlin, New-York: Walter de Gruyter, 173-195.

Guéguen, N. 2002. *Psychologie de la manipulation et de la soumission*. Paris: Dunod

Habermas, J. 1969. *Technik und Wissenschaft als ideologie*. Frankfurt: Suhrkamp.

Hackett-Renner, C. 2004. "Validity effect". In Pohl, R. (ed). *Cognitive illusions. A handbook on fallacies and biases in thinking, judgement and memory*. Hove & New York: Psychology Press, 201-213.

Hamblin, C. 1970. *Fallacies*. London: Methuen.

Harris, P. Pasquini, E., Corriveau, K., Koenig, M. & Clément, F. forth. "The emergence of selective trust in early childhood". In Proust & J. Dokic (eds.). *From Metacognition to Self-Awareness*. Oxford University Press.

Hart, C. 2010. *Critical Discourse Analysis and Cognitive Science: New Perspectives on Immigration Discourse*. Basingstoke: Palgrave.

Horn, L. 1984. "Towards a new taxonomy for pragmatic inference: Q-based and R-based implicature". In Deborah Schiffrin (ed.) *Meaning, Form and Use in Context: Linguistic Applications (GURT '84)*, 11-42. Washington: Georgetown University Press.

Hurford, J. 2007. *The origins of meaning*. Oxford: Oxford University Press.

Hutchby, I. & Wooffitt, R. 1998. *Conversation analysis: The study of talk in interaction*. London: Sage.

Huth, M. & Ryan, M. 2004. *Logic in Computer Science: Modelling and Reasoning about Systems*. Cambridge: Cambridge University Press.

Iacona, A. & Marconi, D. 2005. "Petitio principii: What's wrong?". *Facta Philosophica* 7: 19-34.

Ietcu-Fairclough, I. 2008. "Legitimation and strategic maneuvering in the political field". *Argumentation* 22: 399-417.

Jackson, S. 1996. "Fallacies and heuristics". In: J. van Benthem, F.H. van Eemeren, R. Grootendorst & F. Veltman (Eds.), *Logic and argumentation*. Amsterdam: Royal Netherlands Academy of Arts and Sciences, 101-114.

Jacobs, S., Dawson, E. & Brashers, D. 1996. "Information manipulation theory: a replication and assessment". *Communication monographs* 63(1): 70-82.

Johnson, R. H, 1996. *The Rise of Informal Logic*. Newport News: Vale Press.

Johnson, R. H. & Blair, J. A. 2000. "Informal logic: An overview". *Informal Logic* 20(2): 93-99.

Joule, R.-V. & Beauvois, J.-L. 1998. *La soumission librement consentie*. Paris : Presses Universitaires de France

Kahneman, D. & Tversky, A. 1972. "Subjective probability: a judgement of representativeness". *Cognitive Psychology* 3: 430-454.

Kahneman, D. & Tversky, A. 1973. "On the psychology of prediction". *Psychological Review* 80: 237-251.

Kress, G. 1985. *Linguistic processes in sociocultural practice*. Victoria: Deakin University Press.

Lakatos, I. 1978. *The methodology of scientific research programmes*. Cambridge: Cambridge University Press.

Lakoff, G. & Johnson, M. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.

Levinson, S. 1983. *Pragmatics*. Cambridge: Cambridge University Press.

Levinson, S. 2000. *Presumptive meanings*. Cambridge, MA: MIT Press.

Lewis, D.K. 1969: *Conventions*. Cambridge: Harvard University Press.

Loftus, E. F. 1975. "Leading questions and the eyewitness report". *Cognitive Psychology* 7: 560-572.

Loftus, E.F. & Palmer, J.C. 1974. "Reconstruction of automobile destruction: an example of the interaction between language and memory". *Journal of Verbal Learning and Verbal Behavior* 13: 585-589.

Loftus, E.F. 1993. "Made in memory: Distortions in memory after misleading communications". In G. Bower (ed.). *The psychology of learning and motivation: Advances in research and theory* 30, 187-215. San Diego: Academic Press.

Lorenzen, P. 1969. *Normative logic and ethics*. Hochschultaschenbücher 236. Mannheim: Bibliographisches Institut.

Lorenzen, P. 1987. *Lehrbuch der konstruktiven Wissenschaftstheorie [Textbook of constructive theory of science]*. Mannheim/Vienna/Zurich: Bibliographisches Institut Wissenschaftsverlag.

Lumsden, D. 2008. "Kinds of conversational cooperation". *Journal of Pragmatics* 40: 1896-1908.

MacPherson, B. 1999. "Three misrepresentations of logic". *Informal Logic* 19(3&4): 185-199.

Maillat, D. 2006. "Is Manipulation Relevant?". Invited talk at Universita della Svizzera Italiana, Lugano.

Maillat, D. 2010. "Argumentative illusions and shortcuts: The cognitive underpinnings of manipulation". Paper delivered at the *7th International Conference on Argumentation of the International Society for the Study of Argumentation*, Amsterdam, Netherlands, June 29-July 2.

Maillat, D. & Oswald, S. (forth.). "Constraining context: a pragmatic account of cognitive manipulation". In Hart, C. (ed.). *Critical discourse studies in context and cognition*. Amsterdam: John Benjamins.

Maillat, D. & Oswald, S. 2009. "Defining manipulative discourse: the pragmatics of cognitive illusions". *International Review of Pragmatics* 1 (2): 348-370.

Martinez-Conde, S & Macknik, S. 2008. "Magic and the brain". *Scientific American* (December issue).

Martinich, A. 1984. *Communication and reference*. Berlin, New York: De Gruyter.

Mautner, G. 2005. "Time to get wired: Using web-based corpora in critical discourse analysis". *Discourse & Society* 16 (6): 809-828.

Mautner, G. 2008. "Analysing newspapers, magazines and other print media". In: Wodak. R. & Krzyzanowski, M. (eds) *Qualitative Discourse Analysis in the Social Sciences*. Basingstoke: Palgrave-Macmillan, 30-53.

McCornack, S.A, Levine, T.R., Torres, H.I, Campbell, D.M. 1992. "When the alteration of information is viewed as deception: An empirical test of information manipulation theory". *Communication Monographs* 59: 17–29.

McCornack, S.A. 1992. "Information Manipulation Theory". *Communication Monographs* 59: 1–16.

McGuire, W.J. 1985. "Attitudes and attitude change". In Lindzey, G. & Aronson, E. (eds). *Handbook of social psychology*. (3rd ed., Vol. 2). New York: Random House, 233-346.

Mercier, H. & Sperber, D. 2009. "Intuitive and reflective inferences". In Evans, J. S. B. T. & Frankish, K. (eds) *In Two Minds*. New York: Oxford University Press, 149-170.

Mercier, H. & Sperber, D. forth. "Why do humans reason? Arguments for an argumentative theory". *Behavioral and Brain Science*.

Mercier, H. submitted. "Looking for arguments".

Metts. S. "An exploratory investigation of deception in close relationships: a consequence of violated expectations". *Journal of social and personal relationships* 6: 159-179.

Milgram, S. 2004 [1974]. *Obedience to Authority: An Experimental View*. New York: Harper Collins.

Millikan, R. 1998. "A common structure for concepts of individuals, stuffs, and real kinds: More Mamma, more milk and more mouse". *Behavioral and Brain Sciences* 21: 55-100.

Mitnick, K. 2003. *The art of deception. Controlling the human element of security*. Indianapolis: Wiley Publishing.

Moeschler, J. 2009. "Pragmatics, propositional and non-propositional effects: can a theory of utterance interpretation account for emotions in verbal communication?". *Social Science Information* 48 (3) (Special Issue *The Language of emotion: conceptual and cultural issues*): 449-466.

Morency, P., Oswald, S. & Saussure, L. de. 2008. "Explicitness, implicitness and commitment attribution: a cognitive pragmatic approach". *Belgian Journal of Linguistics* 22: 197-219.

Moshman, D., & Geil, M. 1998. "Collaborative reasoning: Evidence for collective rationality". *Thinking and Reasoning* 4(3): 231-248.

Musolff, A. 2004. *Metaphor and political discourse: Analogical reasoning in debates about Europe*. Basingstoke: Palgrave Macmillan.

Nissani, M. 1990. "A cognitive reinterpretation of Stanley Milgram's observations on obedience to authority." *American Psychologist* 45: 1384-1385.

Noveck, I. 2001. "When children are more logical than adults: Investigations of scalar implicature". *Cognition* 78(2): 165-188.

Noveck, I.A. and Sperber, D. (eds). 2004. *Experimental Pragmatics*. Basingstoke: Palgrave Macmillan.

O'Halloran, K. 2003. *Critical discourse analysis and language cognition*. Edinburgh: Edinburgh University Press.

O'Keefe, D. 1996. "Argumentation studies and dual-process models of persuasion ". In van Benthem, J., van Eemeren, F. H., Grootendorst, R., & Veltman, F. (Eds). *Logic and argumentation*. Amsterdam: North-Holland, 61-76.

O'Keefe, D. 2002. *Persuasion: Theory and research* (2nd ed.). Thousand Oaks, CA: Sage.

Origgi, G. & Sperber, D. 2000. "Evolution, communication and the proper function of language". In P. Carruthers & A. Chamberlain (Eds), *Evolution and the Human Mind: Language, Modularity and Social Cognition*. Cambridge: Cambridge University Press, 140-169.

Oswald, S. 2007a. "Argumentation and Cognition: Can Pragma-Dialectics Interplay with Pragma-Semantics?". *CADAAD (Critical Approaches to Discourse Analysis Across Disciplines) eJournal* 1, 148-165.

Oswald S. 2007b. "Towards an Interface between Pragma-Dialectics and Relevance Theory". *Pragmatics and Cognition* 15(1): 179-201.

Oswald, S. 2010. "Relevance for effectiveness: cognitive underpinnings of information-processing". Paper delivered at the *7th International Conference on Argumentation of the International Society for the Study of Argumentation*, Amsterdam, Netherlands, June 29-July 2. (accessible online: http://docs.google.com/uc?export=download&id=0B99SoJAZNQsZNjgwMWQ0MGYtYjlmYi00M2E3LWE5YTItNWZhOWI1MGVjMjE4&authkey=CP_S8dUE)

Oswald, S. & Rihs, A. 2010. "Metaphors and consent in political discourse: a pragmatic view on some perlocutionary issues". Paper delivered at the International Conference *CADAAD 2010*, Lodz, Poland, September 13.

Oswald, S. & Maillat, D. 2010. "The cognitive determinants of manipulation". Paper delivered at the International Conference *Utterance Interpretation and Cognitive Models III* (*UICM3*), Brussels, February 5-7.

Paicheler, G. 1985. *Psychologie des influences sociales. Contraindre, convaincre, persuader*. Neuchâtel: Delachaux & Niestlé.

Park, H. & Reder, L. 1991. "Locus of the Moses illusion: imperfect encoding, retrieval, or match?". *Journal of memory and language* 30: 385-406.

Park, H. & Reder, L. 2004. "Moses illusion". In Pohl, R. (ed). *Cognitive illusions. A handbook on fallacies and biases in thinking, judgement and memory*. Hove & New York: Psychology Press, 275-291.

Parret, H. 1978. "Eléments d'une analyse philosophique de la manipulation et du mensonge. Prépublication." *Documents de Travail et Prépublications. Centro Internazionale di Semiotica e di Linguistica. Universita di Urbino* 70, 1-33.

Pavlidou, T. 1991. "Cooperation and the choice of linguistic means: some evidence from the use of the subjunctive in Modern Greek". *Journal of Pragmatics* 15: 11-42.

Perelman, C. & Olbrechts-Tyteca, L. 1969. *The New Rhetoric: A treatise on argumentation*. Notre Dame: University of Notre Dame Press.

Perelman, C. & Olbrechts-Tyteca, L. 2008 [1958]. *La nouvelle rhétorique: Traité de l'argumentation*. Bruxelles: Editions de l'Université Libre de Bruxelles.

Perry, J. 1986. "Thought without representation". *Supplementary Proceedings of the Aritotelian Society* 60: 263-283.

Pickrell, J., Bernstein, D. & Loftus, E. 2004. "Misinformation effect". In Pohl, R. (ed). *Cognitive illusions. A handbook on fallacies and biases in thinking, judgement and memory*. Hove & New York: Psychology Press, 345-361.

Pohl, R. (ed). 2004. *Cognitive illusions. A handbook on fallacies and biases in thinking, judgement and memory*. Hove & New York: Psychology Press.

Pohl, R. 2010. "The domain of suggestionality: Communicative and other influences on cognitive illusions". Conference given at the University of Neuchâtel, Switzerland, 11 May 2010. Slides available on http://www2.unine.ch/webdav/site/grsp/shared/documents/DocsPresentations/PohlNeuchatel.pdf)

Polyzou, A. in prep. "(Ideological) knowledge management and gender in Greek lifestyle magazines: a socio-cognitive discourse analytical approach". PhD dissertation, Lancaster University.

Popper, K. 2009 [1959]. *The logic of scientific discovery*. New-York: Routledge.

Popper, K. 2009 [1963]. *Conjectures and refutations*. New-York: Routledge.

Puzynina, J. 1992. *Jezyk w swiecie wartosci* [*Language in the world of values*]. Warsaw, Poland: Panstwowe Widawnictwo Naukowe.

Reber, R. 2004. "Availability". In Pohl, R. (ed). *Cognitive illusions. A handbook on fallacies and biases in thinking, judgement and memory*. Hove & New York: Psychology Press, 147-163.

Reboul, A. 2002. "Semantic transparency, semantic opacity, states of affairs, mental states and speech acts". In Anolli, L. & Ceciri, R. (eds), *Say not to say. New perspectives on miscommunication*. Amsterdam: IOS Press, 46-71.

Récanati, F. 1979. "Insinuation et sous-entendu". *Communications* 30: 95-106.

Reid, T. 1764: *Inquiry into the Human Mind*.

Reid, T. 1983 [1785]. *Essays on the intellectual power of man*. In *Philosophical works*, Hildesheim, Georg Olms Verlag.

Reisigl, M. 2010. "An ideological journey. On procedures for controlling interaction in public places." Plenary lecture delivered at the international conference CADAAD 2010, Lodz, Poland, 13 September.

Reisigl, M. & Wodak, R. 2003. "The discourse-historical approach (DHA)". In Wodak, R & Meyer, M. (eds). *Methods for Critical Discourse Analysis*. London: Sage, 87-121.

Rigotti, E. 2005. "Towards a typology of manipulative processes". In Saussure, L. de & Schulz, P. (eds), *Manipulation and Ideologies in the Twentieth Century: Discourse, language, mind*. Amsterdam: Benjamins, 61-83.

Saussure, L. de. 2005. "Manipulation and cognitive pragmatics: Preliminary hypotheses". In Saussure, L. de & Schulz, P. (eds). *Manipulation and Ideologies in the Twentieth Century: Discourse, language, mind*. Amsterdam: Benjamins, 113-146.

Saussure, L. de. 2007. "Procedural pragmatics and the study of discourse". In Saussure, L. de & Schulz, P. *Pragmatic interfaces*. Special issue of *Pragmatics & Cognition* 15 (1): 139-159.

Saussure, L. de & Oswald, S. 2008. "L'engagement comme notion cognitive associée au destinataire". *L'analisi linguistica e letteraria* XVI (Special issue: Word meaning in argumentative dialogue), 475-488.

Saussure, L. de & Oswald, S. forth. "Argumentation et engagement du locuteur. Pour un point de vue subjectiviste". *Nouveaux Cahiers de Linguistique Française*.

Saussure, L. de & Schulz, P. (eds). 2005. "Introduction". In Saussure, L. de & Schulz, P. (eds). *Manipulation and Ideologies in the Twentieth Century: Discourse, language, mind*. Amsterdam: Benjamins, 1-14.

Saussure, L. de & Schulz, P. (eds). 2005. *Manipulation and Ideologies in the Twentieth Century: Discourse, language, mind*. Amsterdam: Benjamins.

Schegloff, E. (1999b). 'Naivete vs sophistication or discipline vs self-indulgence: a rejoinder to Billig', in Discourse and Society 10 (4): 577–8.

Schegloff, E. 1999a. "Schegloff's texts as Billig's data: A critical reply". *Discourse & Society* 10 (4): 558–72.

Schein, E. 1956. "Some Observations on Chinese Methods of Handling Prisoners of War". *The Public Opinion Quarterly* 20(1) (Special Issue on Studies in Political Communication), 321-327.

Searle, J. R. 1969. *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.

Searle, J. R. & Vanderveken, D. 1985. *Foundations of illocutionary logic*. Cambridge: Cambridge University Press.

Skinner, B. F. 1953. *Science and Human Behavior*. New York: Macmillan.

Sokal, A. & Bricmont, J. 1997. *Impostures intellectuelles*. Paris: Odile Jacob.

Sperber, D. 2000a. "Metarepresentations in an evolutionary perspective". In Dan Sperber (ed). *Metarepresentations: A Multidisciplinary Perspective*. Oxford: Oxford University Press: 117-137.

Sperber, D. 2000b. *Metarepresentations: A Multidisciplinary Perspective*. Oxford: Oxford University Press.

Sperber, D. forth. "A naturalistic ontology for mechanistic explanations of the social sciences". In Demeulenaere, P. (ed). *Analytical sociology and social mechanisms*. Cambridge: Cambridge University Press.

Sperber, D. & Mercier, H. In press. "Reasoning as a social competence". In Elster, J. and Landemore, H. (eds) *Collective Wisdom*.

Sperber, D. & Wilson, D. 1995 [1986]. *Relevance. Communication and cognition* (2nd edition). Oxford: Blackwell.

Sperber, D. & Wilson, D. 1997. "The mapping between the mental and the public lexicon". *UCL Working Papers in Linguistics* 9: 1-20.

Sperber, D. Cara, F. & Girotto, V. 1995. "Relevance Theory explains the selection task". *Cognition* 57: 31-95.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G. & Wilson, D. 2010. "Epistemic vigilance". *Mind and Language* 25(4): 359-393.

Strawson, P. 1964. "Intention and convention in speech acts". *Philosophical Review* 73(4): 439-460.

Swartz, N. 1997. "The Concepts of Necessary Conditions and Sufficient Conditions". Last accessed 02.09.2008 on http://www.sfu.ca/philosophy/swartz/conditions1.htm.

Taillard, M.-O. 2000. "Persuasive communication: the case of marketing". *University College London Working Papers in Linguistics* 12: 145-172.

Tanaka, K. 1994. *Advertising Language. A pragmatic approach to advertisements in Britain and Japan*. London – New York: Routledge.

Tendahl, M. 2009. *A hybrid theory of metaphor: Relevance Theory and Cognitive Linguistics*. Basingstoke: Palgrave Macmillan.

Tendahl, M. & Gibbs, R. 2008. "Complementary perspectives on metaphor: Cognitive Linguistics and Relevance Theory". *Journal of Pragmatics* 40 (11): 1823-1864.

Thagard, P. 2010. "Cognitive science". *Stanford Encyclopedia of Philosophy*. Online version, available at http://plato.stanford.edu/entries/cognitive-science/.

Thomas, J. 1995. *Meaning in interaction: an introduction to pragmatics*. Harlow: Longman.

Toulmin, S. 2008 [1958]. *The uses of argument*. Cambridge, New York: Cambridge University Press.

Trew, T. 1979. "Theory and ideology at work". In Fowler, R. *et al*. *Language and Control*. London: Routledge and Kegan Paul, 94-116.

Trivers, R. 1971. "The evolution of reciprocal altruism". *Quarterly Review of Biology* 46: 35-57.

Turner, R. E., Edgley, C. & Olmstead, G. 1975. "Information control in conversations: honesty is not always the best policiy". *Kansas Journal of Sociology* 11: 69-89.

Tversky, A. & Kahneman, D. 1974. "Judgment under Uncertainty: Heuristics and Biases". *Science* 185(4157): 1124-1131.

Tversky, A. & Kahneman, D. 1983. "Extensional versus intuitive reasoning: the conjunction fallacy in probability judgement". *Psychological Review* 90: 293-315.

Walton, D.N. 1985. *Arguer's Position: A Pragmatic Study of* Ad Hominem *Attack, Criticism, Refutation, and Fallacy*. Westport, CT: Greenwood Press.

Walton, D. 1987. *Informal fallacies. Towards a theory of argument criticisms*. Pragmatics and Beyond Companion Series 4. Amsterdam: John Benjamins.

Walton, D. 2006. *Fundamentals of critical argumentation*. New-York: Cambridge University Press.

Walton, D. 2008. *Informal Logic*. Cambridge: Cambridge University Press.

Wason, P. 1960. "On the failure to eliminate hypotheses in a conceptual task". *The Quarterly Journal of Experimental Psychology* 12(3): 129-140.

Watson, J. 1913. "Psychology as a behaviorist views it". *Psychological Review* 20: 158-177.

Weiss, G. & Wodak, R. (eds.) 2003. *Critical Discourse Analysis: Theory and Interdisciplinarity*. Basingstoke: Palgrave Macmillan.

Whately, R. 1843. *Elements of Logic*. Boston: James Munroe and Company. (via Google Books: http://books.google.com/books?id=XnUBAAAAYAAJ&oe=UTF-8, last accessed 14.04.2010)

Whiten, A. & Byrne, R. 1997. *Machiavellian Intelligence II: Extensions and Evaluations*. Cambridge: Cambridge University Press.

Widdowson, H. 1995a. "Discourse analysis: a critical view". *Language and Literature* 4(3) 157-172

Widdowson, H. 1995b. "Review of *Discourse and social change*". *Applied Linguistics* 16 (4): 510-516.

Widdowson, H. 1996. "Reply to Fairclough. Discourse and interpretation: conjectures and refutations". *Language and Literature* 5(1): 57-69

Widdowson, H. 1998. "The theory and practice of Critical Discourse Analysis". *Applied linguistics* 19(1): 136-151

Widdowson, H. 2000. "On the limitations of linguistics applied". *Applied Linguistics* 21(1): 3-25

Wilson, D. & Sperber, D. 1981. "On Grice's theory of conversation". In P. Werth (ed), *Conversation and discourse*. London, Croom Helm, 155-198.

Wilson, D. 1994. "Relevance and understanding". In in G. Brown, K. Malmkjaer, A. Pollitt & J. Williams (eds), *Language and understanding*. Oxford: Oxford University Press: 35-58.

Wilson, D. 2000. "Metarepresentation in Linguistic Communication". In D. Sperber (ed), *Metarepresentations: A Multidisciplinary Perspective*. Oxford: Oxford University Press, 411-448.

Wilson, D. 2003. "New Directions for Research in Pragmatics and Modularity". *UCL Working Papers in Linguistics* 15: 105-127.

Wodak, R. & Meyer, M. (eds). 2001. *Methods of Critical Discourse Analysis*. London: Sage.

Wodak R. & Meyer, M. (eds). 2009. *Methods of Critical Discourse Analysis* (2nd edition). London: Sage.

Wodak, R. & Reisigl, M. 2009. "The Discourse-Historical Approach (DHA)". In Wodak, R. & Meyer, M. (eds). *Methods of Critical Discourse Analysis* (2nd edition). London: Sage, 87-121.

Wodak, R. 2000. "Recontextualisation and the transformation of meaning: a critical discourse analysis of decision making in EU-meetings about employment policies". In S. Sarangi & M. Coulthard (eds.). *Discourse and Social Life*. London: Pearson Education, 185-206.

Wodak, R. 2006. "Critical Linguistics and Critical Discourse Analysis". In J.-O. Östman and J. Verschueren (eds) *Handbook of Pragmatics*, Amsterdam: John Benjamins, 1-24.

Wodak, R. 2007. "Pragmatics and Critical Discourse Analysis: a cross-disciplinary inquiry". *Pragmatics & Cognition* 15 (1): 203-225.

Woods, J. & Walton, D. 1989. *Fallacies: Selected Papers, 1972-1982*. Dordrecht/Providence: Foris Publications, PDA 9.

Woods, J. 1995. "Fearful symmetry". In H. V. Hansen & R.C. Pinto (eds). *Fallacies: Classical background and contemporary developments*. University Park, PA: The Pennsylvania State University Press.

Yzerbyt, V. & Corneille, O. (eds.) 1994. *La persuasion*. Neuchâtel: Delachaux & Niestlé.

Zajonc, R. 1968. "Attitudinal effects of mere exposure". *Journal of Personality and Social Psychology. Monograph supplement* 9(2): 1-27.