The relevance-affective model: explaining narrative empathy within relevance theory

Ismaël Pozner

Supervisor: Louis de Saussure
University of Neuchâtel
Master in cognitive science / Spring 2022

Friday, 29 July 2022

Table of content

1. Introduction 3

2. Empathy in Appraisal Theory 6

3. Narrative Empathy 15

3.1. The role of engagement 20

4. Relevance Theory 24

5. The model of narrative empathy in Relevance theory 26

5.1. Procedural meaning and affective effects 27

5.2. Metacognitive acquaintance and perspective-taking 32

5.3. Engagement as product of goal relevance 38

6. Discussion with other theoretic models 43

6.1. Representation Matching 44

6.2. The Simulation Theory 45

6.3. The Perception-Action model 48

6.4. Perceptual symbol systems and non-propositionality 54

6.5. Emotion as the inferential output 59

7. Conclusion 60

8. Acknowledgments: 63

References: 64

Abstract:

Narrative empathy is the sharing of similar feelings with a fictional character as a result of perceiving or imagining her depicted situation. The notion originating from cognitive narratology is involved in both appraisal theories of emotion and pragmatic theories of relevance. However, despite recent attempts to bridge cognition and affective states, no relevance-theoretic study has directly dealt with empathy nor narrative empathy. The purpose of this work is to present a model of narrative empathy within relevance theory. Firstly, we will see how appraisal theories explain empathy in terms of appraisal matching, and its prerequisites. Secondly, we will examine the way cognitive narratology handles narrative empathy and mentalistic inferences. We will see how it highlights the role of engagement in narrative emotions. We will adapt their notions to fit the relevance-theoretic framework. Thirdly, after explaining the main principles of relevance theory, we will build a model of narrative empathy, called the Model of Affective Relevance, using previous insights and notions developed by the field of relevance theory, successively adding procedural meaning, affective effects, metacognitive acquaintance, and goal relevance. We describe narrative empathy as a continuous inferential process, involving a constant affective attunement in which processing of emotional cues of the character's situation activates relevant representations of a specific kind called 'metacognitive acquaintance'; it allows mentalistic inferences linking events and characters' states, and implies an update of the reader's affective state. Seemingly matching the target's affective state allows the empathizer through affective procedures a better retrieval of relevant representations. I propose that empathy is one possible outcome of the model in which the empathizer engages her own goals in the target, and there must be connivance between target's attributed goals and representations and the empathizer's, from the latter's perspective.

1. Introduction

Stories allow us to create a privileged perspective on a character's social and mental life. Most emotions we experience in narratives are through and for fictional characters. It is made possible by perspective-taking as well as empathy, the ability to feel others' emotions. As literature creates a sense of intimacy with characters' mental lives, narratives allow us to know more about their feelings than we do with real humans. Although there are many cases in which we do not feel the same depicted emotion as one specific character, there are many other cases where characters act as a proxy to experience the narrative world. We feel through them,

even if they live situations we are unfamiliar with, and we learn through them that we share more with others than previously envisioned (Keen 2006). Stories allow us to create a very privileged perspective on its characters' social life and mental lives, and empathy seems to be the main way to achieve this bond. In the following work, the terms 'empathy' and 'narrative empathy' are employed according to the context of use, the former being for general cases and mostly real people, the latter for narrative objects, essentially fictional characters.

Narrative empathy has been already treated by cognitive narratologists and is generally explained it as a composite of different mechanisms. Oatley (1999b) explains empathy in narratives as emerging from the planning processor, a computational ability mapping mental information to people to predict their behaviour. Later, Oatley and Djikic (2018) explains narrative processing in terms of simulation using mental models. Consequently, narrative empathy would rely on mental models as well. Keen (2010) indicates narrative empathy as not inherently functioning differently from empathy, using perspective-taking, our experiences and memory, and as situated at an intersection between affects and cognition (Keen 2006). This picture of narrative empathy offers many insights on the nature and functioning of empathy, though it is not clear how that works at an inferential level, that is, it does not offer a unified account with other information processing as in relevance theory, or other theories of emotion elicitation. Thus, adapting their insights to other frameworks could provide a finer-grained analysis.

In relevance theory, narrative empathy, and to a larger extent empathy, has never been seriously explored. This lack of research for emotions reflects an old trend inherited from historical approaches. Those approaches consider emotions as entirely separate from cognition, leading to neglect them from theoretical accounts. Indeed, relevance theory has focused on propositional contents, and tried to circumvent the issue of non-propositionality using the notion of impressions (Sperber & Wilson 2015). Impressions are wide arrays of weakly manifest implicatures that remain propositional at core. This has allowed the theory to keep using propositions when faced with feelings, emotions, and other non-paraphrasable and ineffable effects. According to the extant account, their ineffability is simply due to a large array of implicatures too wide to be enumerated. A branch of the relevance-theoretic field was dedicated to study literature, but it was historically focused on stylistics and how certain textual features were interacting with relevance and create cognitive effects (Green 1993; Trotter 1992; Cave & Wilson 2018); emotions were still mostly left out of the picture. Yet, in the last decade, a flourishing area of relevance-theoretic research proposed that expressives and other non-propositional contents bears procedural meaning (Blakemore 2011; Wharton 2016). Indeed, similarly to logical discourse connectors, expressives and other non-verbal behaviours encode procedures, those are described as specific instructions to retrieve a specific

space of representations, this can be done by activating procedure-related cognitive modules, such as mind reading, emotion reading, syntactical modules, etc. Soon, the question of art and literature would follow: Pilkington (2000) and Kolaiti (2015, 2019) with poetic effects, Cave and Wilson (2018) with sensorimotor resonance, de Saussure (2021) with emotionally charged experiences resonance, and Fabb (2021) with his discrepancy-based model of experiences of ineffable significance, all proposing original accounts on how art is processed and ultimately leads to emotions. However, these analyses have focused on art and literature at a global level, and rarely touched specifically on narratives. Even when they did, they did not differentiate categories of narrative emotions as found in cognitive-narrative theories, such as evoked emotions (elicited by remembering memories or experiences) or character-driven emotions (elicited by observing/inferring a character's situation and expression, such as in narrative empathy).

Most of the cited works above were arguably focused on the evoked emotions in literature, but very little has been said about the character-driven ones, bearing a more mentalistic view on emotions. I see two reasons for the dearth of work on emotions in pragmatics: for one, the inclusion of emotions in relevance-theoretical accounts on art is relatively new. For another, emotions and empathy are multifaceted notions, requiring different theoretical backgrounds and mechanisms foreign to relevance theory. To talk about narrative empathy and character-driven emotions, we need a clear explanation of how we appraise others' situations in a narrative context. Alas, there was no real attempt of dealing with appraisals within narratives specifically in pragmatics. But other studies have recently tried to introduce notions into relevance theory from appraisal theory (for example, Wharton et al. (2021)), which is much better equipped to deal with empathy as it is focused on emotions. As a result, we will take full advantage of both theories to handle narrative empathy. I argue that tools developed by the field of relevance can deal with character-driven emotions, and more specifically for this present work, narrative emotions elicited through empathy. With procedural meaning, affective effects, and notions from the appraisal theory, it has become possible to include feelings in the pragmatic equation. Relevance theory would greatly benefit from an account on narrative empathy, as we would have a better understanding of the how appraisal processes function in pragmatics, apply to fictions, but also to people in general. Having empathy in relevance theory is the perfect example of emotions aiding cognition, thus will show how both work together, and will move the *non-propositional* debate forward. Consequently, narrative empathy has not benefited yet from this affective revolution in relevance theory, and we will attempt to remedy this absence.

Throughout this paper, I will try to build a model to account for narrative empathy in a relevance-theoretic framework using recent findings and completing it with empathy-related notions in affective science and cognitive narratology.

The present work is structured as follows. In section one, we will look at an overview of empathy in affective science, how it is apprehended in the appraisal theory, and how situational factors and similarities with the target can influence the empathic process. In section two, I will present how cognitive narratology conceptualises narrative empathy and delimits emotions of empathy among other narrative emotions. We will have a look at the notions of planning processor, fiction as simulation and see how they explain character-driven emotions and thus empathy. In section three, I will present relevance theory and its main principles and functioning, it will constitute the chassis of our model. In section four, I will present the model of affective relevance (MAR) aiming to account for narrative empathy within relevance. We will start from the standard relevance-theoretic model and by successively introducing new relevance-theoretical notions as add-ons to improve each version of the model and come closer to explaining (narrative) empathy. In section five, in a discussion, we will compare the MAR with other existing models and proposals with similar aims, such as the Fabb's (2021) representation matching, the simulation theory (Johnson-Laird 1983), the perception-action model (Preston & de Waal 2002), perceptual symbols (Barsalou 1999), and Piskorska's (2018) proposal of emotion as information and look at how their conclusions matches or even complements our model. This cross-examination will shed light on the topic of conceiving thoroughly non-propositional contents within relevance theory. After consecutively providing new add-ons to the MAR, the final version of the model will be shown to explain narrative empathy and will provide three main conclusions on the nature of empathy and affects in inferential processes. Firstly, empathy not only functions with the same inference-making mechanisms as in relevance theory, but it also relies on a special kind of representation about our experience of one's and others' mental lives for mentalistic inferences, automatically activated from perception of behavioural cues, called metacognitive acquaintance. Secondly, that empathy and narrative empathy function virtually the same, what distinguishes them is a matter of affective goals. Thirdly, the MAR suggests that cognitive effects, as described by Sperber and Wilson (1995), are not the only outcome of inferential processes, as it can result in a change of affective state, devoid of propositional content, supporting de Saussure and Wharton (2020).

2. Empathy in Appraisal Theory

Empathy is the spontaneous sharing of an affect from witnessing, hearing, or reading about it in another's emotional state (Keen 2006). In that sense it is a vicarious experience. The word "empathy" comes in English as translation of the German word "Einfühlung". It originally meant "feeling one's way into" an artwork or another person, it was then then adapted to describe the reading experience in terms of our "tendency to feel ourselves into what we perceive or imagine" (Titchener 1909). Lipps (1903) was the first to introduce "Einfühlung" as a mechanism where perception of another's gesture result in feeling the same emotion in the perceiver without proposing any underlying cognitive process for it.

The study of empathy lies between the domains of cognitive and affective science. It involves both feeling and thinking (Keen 2006). Different cognitive processes play a role in empathy, such as memory, experiences, and perspective-taking (Keen 2010). In the last decades, a growing branch of research advances that emotions and rationality are not as dichotomous as previously thought (Damasio 1991; Cosmides & Tooby 1992; Wharton & Strey 2019). Empathy (or affective empathy) is to be distinguished from perspective-taking, (sometimes called cognitive empathy), which is the ability to estimate others' thought or feelings through mind-reading (Batson 1991) and temporarily adopt the characters' goals and plans (Aarts, Gollwitzer & Hassin 2004). Unlike empathy, it does not require the observer to *feel*, but merely to *know* the target's states. However, perspective-taking is commonly referred as a necessary component of empathy (Keen 2015; Wondra & Ellsworth 2015). Thus, empathy requires mentalistic inferences, as Keen (2015) describes it as follows: "the empath who engages in perspective-taking employs observation of the other and knowledge of that person" and goes on "empathy in this sense is a more obviously cognitive operation that depends on having a theory of (another's) mind (ToM)." (2015: 131). In the same vein, Coplan (2004) envisions empathy "as a complex imaginative process involving both emotion and cognition" (2004: 143). It appears that empathy, despite being tied to emotions and affects, strongly involves cognitive-inferential components, and thus uses mechanisms from both sides.

I also distinguish empathy (feeling the same emotion as that of another from observing it) from sympathy (feeling a different emotion from observing another's). Sympathy refers to a mismatch with the target's feelings, encompassing cases for "feeling for someone" and compassion (Keen 2014), whereas empathy is better encompassed with feeling with/as someone and "feelings *like* the others' feelings" (original emphasis) (Keen 2015: 132). For example, a mother feeling sad for her child coming back from school crying because of conflicts with her schoolmates is empathy, whereas the mother feeling sorry for her child is sympathy.

As a disclaimer, we will use the term "target" throughout this work to indicate the person with which the observer will (potentially) empathize. Contrary to what it might imply, the term *target* does not mean that the observer unavoidably *intends* or *wants* to empathize with the empathized person, the empathic process is mainly automatic and may not necessarily require any deliberate specific reflective thought from the empathizer to work.

The appraisal theory of emotions describes first-hand emotions as resulting from the evaluation of a situation by the observer (Lazarus 1991*;* Wondra & Ellsworth 2015). The theory postulates three claims about emotions, firstly, that emotions are appraisals of situations. Secondly, emotions of different qualities occur along a continuum with no clear boundaries. Thirdly, emotions have universal patterns of appraisals. Emotional experience is thus made of automatic evaluations of situation, which are appraisals. Following that, various dimensions of appraisals have been found and proposed (see Wondra & Ellsworth (2015), and Clore & Orthony (2008)), bringing suggestions of variables with which we would appraise situations, such as the pleasantness of a situation, the amount of effort to deal with a situation, the situational control over the situation, the responsibility agents have for the situation, or how much attention is drawn to the situation or diverted (Smith & Ellsworth 1885), or the certainty about what is happening or what will come next in a situation (Ellsworth & Sherer 2003). For example, an unpleasant, certain, a low-situational control and high other-agency appraisal would correspond to anger, but if you increase the situational control to high, it will most likely become sadness (Wondra & Ellsworth 2015).

Event appraisal is intimately linked to goal congruence. Lazarus (1991) explains that goal-congruence is a crucial factor when it comes to emotional valence: positive emotions stem from goal congruent events whereas negative emotions stem from goal-incongruent events. In other words, the dimensions of appraisal presented above could be considered as being derived from goal-related appraisals, that is, how much an event is evaluated to hinder, benefit, make it uncertain, leave us time, or give us control over the achievement of one of our goals will affect the valence, and intrinsic pleasantness of the emotion. Wondra and Ellsworth (2015) suggest that it implies that in empathic responses, if the target's situation is not relevant to the observer's goals, no emotions should be elicited (though intrinsic pleasantness of the situation can still influence the observer's appraisal). Conversely, if the target's situation obstructs or helps progress observer's goals, or that the target's wellbeing is tied to those goals, then her appraisal of goal-congruence should dictate the valence of the emotion. The authors thus remarked that for an appraisal of a target's situation to occur, it requires that the observer attends the target's situation, that she has enough information to appraise the other's situation, and that she evaluates that situation as good or bad relative to the progression of her goals. It

follows that empathy as well must be a matter of goal congruence, since in cases of empathic responses, the observer appraises the target's situation.

Wondra and Ellsworth (2015), using the appraisal theory, created a model vicarious experiences defining empathy as one outcome of the appraising process. That is, when the appraisal of an individual matches that of another target individual, and as a result of perceiving the target's emotional state. They proposed a continuous line between non-matching vicarious emotions and empathic emotions: the closer the observer's appraisal is to the target's, the closer we are to an empathic response. In sum, there are no strict boundaries between the two. The alternative case, when it does not match, is common. It might be that the observer feels an emotion in response to seeing a vicarious emotion, but a different one due to them having different information or different psychological states at the time of the appraisal. Consequently, these factors make the appraisals diverge from that of the target. Sympathy, being non-matching by nature, is comprised by this case. Another case is that the observer does not feel any emotion in response to vicarious emotions, because her appraisal is neutral or because of a lack of information, or unfamiliarity due to novelty. So, the observer does not have the information necessary to appraise the situation. The empathic process largely relies on the observer's past experiences and can be activated through direct or indirect perception. They also presented a memorial categorisation in which first-hand emotions (events directly experienced in the past by the reader) are distinguished from second-hand emotions (emotional cues perceived by the reader in others experiencing the event, or emotions inferred from the telling of the event by a third party). These two ways of emotionally marking experiences help form an adapted appraisal in the right situation. Wondra and Ellsworth underlined that *empathic failures* do not really exist, as they imply that the default outcome is an empathic response, whereas this view suggests that "matching is not an inherent feature of the process" (2015: 415). A criticism to this view is that there is no way for the empathizer to know what the target exactly felt like or what her appraisal was in a given situation, feelings are inaccessible to anyone but the target. So, the observer necessarily matches with emotional cues, situational and non-verbal, to infer what the target is currently mentally experiencing, based on her own past experiences, shared experiences with the target, and general understanding of emotional cues correlating with specific emotional/mental states in others. In summary, Wondra and Ellsworth (2015) showed the relationship between appraisal and empathy, and it implies that empathy requires a certain type of knowledge to link specific cues to related specific mental experiences, acquired from first-hand and second-hand experiences to interpret others' emotions, and in that sense, empathy is fundamentally inferential and "might not be very different from the other vicarious emotional experiences" (2015: 415).

From this point on, we will keep this perspective in which the empathic observer does not *truly* match what the target is experiencing, but actually matches with what she *thinks* the target is experiencing. This distinction is crucial, as there is no way to know for sure what others feel beside correlations between cues from their situation/their behaviours and assumed mental states. So, in this paper, what empathy comes down to is essentially the process of inferring what the target feels: there is no direct matching to someone else's mental states. I maintain that it can be true that we are empathic with a target but in reality, we are not actually matching with her *actual* experience of the situation, just because what we actually match our mental state to is *our attribution* of what the target is experiencing. The advantage of that view is that we can then easily explain why we can empathise with narrative characters who do not truly experience anything since they are fictional: we match with the inferred mental states, through description of their emotional cues and their current situation, as we do with real people. This can happen without any conscious effort, as empathy was proposed to take a faster route, especially when attending certain facial cues, or information from the sensory cortex directly transferred to the reflective cortex. This process might facilitate empathy in certain situations and automatically activate representations of that same state in the observer (Preston & de Waal 2002). If empathy is not an objective state matching with a target, we need to conceive empathy differently.

To allow us to make predictions about empathy without requiring the target and the observer to affectively match, we will take a narrower definition of empathy that of Vignemont and Singer (2006). They describe empathy as the case when:

1. An observer is in an affective state.
2. The observer's affective state is isomorphic to the target's affective state.
3. The observer's affective state is elicited from observing or imagining a target's affective state
4. The observer knows that the source of her affective state is the target's state.

In (1), it is obviously a requirement for empathy. In (2), we will consider *isomorphic* here as the observer attribution that her affective state matches with that of the target. It is because according to our view, empathy is not a state of affairs, but an assumption of the observer. Our conception of empathy considers that the match may never occur *effectively*, though the observer can attribute that there is a match between hers and that of the target. In other words, it can *subjectively* match from the observer's point of view. In (3), it implies that the observer infers the target's state even this one is not present or fictional. Cues of the target's situations need to be directly perceived by senses; they can be inferred. Thus, we can *simulate* events and others affective states in their absence, as for example, during reading. In (4), the observer

requires to distinguish between her own affective state and that of the target to in turn attribute a match between both. This condition was also proposed by Coplan (2004), who states that this self-other differentiation is necessary to distinguish empathy from emotional contagion. She also points out that it also allows the observer to simulate the target's mental while the observer experiences her own mental state. Maintaining a separate line between the two makes comparing them possible, and thus makes empathy possible. With this definition of empathy, we can study it from the observer's point of view without making any assumptions about the actual target's affective state. That last point is crucial when we want to describe a target that has no mental state in actuality (e.g., fictional characters). In other words, this definition allows for narrative empathy to be described to work functionally the same as regular empathy.

An implication of that view is that we can empathise with *erroneous* attribution of the target's experience, and it still counts as empathy. It is because "erroneous" here might just designate the normal case rather than a dysfunctional outcome, or just to be put on a continuum, because we arguably never feel the same as the target. The so-called *narrative empathic inaccuracies* in literary works (Keen 2006, 2015), that is when the audience does not empathize in passages intended by the author, or empathises when unintended, producing conclusions unintended by the author, is just a question of perspective and reflects the differences in attributing experiences across individuals (due to different backgrounds, preferences experiences among readers), not necessarily a flawed empathic process. What gives the observer the impression of having a matching state is because the attributed, inferred and felt state yields predictions consistent with other observed behavioural cues. So, her attribution receives support from inferences that are confirmed. Confirmations in turn solidify the degree of confidence in this attribution for later processes. Since the empathisers cannot know for sure what the target feels, she relies on indirect cues, her past experiences of similar situations, and models of others' mental lives to essentially 'guess' (it is essentially *inferring*) the target's current experience. Surely in practice, the more the target and the observer share common experiences and representations, the more likely their actual states might match if measuring it was possible, and thus increasing the likelihood of an empathic response, but it does not totally brush off the possibility of empathic inaccuracies or maybe attaining an empathic response despite mistaken assumptions. The main asset of the empathic process is that it offers the possibility of reverse-engineering the target's experience through constructing oneself a similar emotional state and feeling it. Subsequently, feeling a state similar to that of the target arguably helps the observer getting closer to knowing what the target experiences and understand her.

A range of cognitive mechanisms is thought to enable and contribute to cases of empathy. In Hoffman (2000), state-matching between an observer and an individual operates through two paths, a fast one, also called primitive empathy. This more primitive route includes mimicry,

conditioning, direct association as mechanisms, and a slower one involving more advanced functions with mediated association[1] and perspective-taking. The fast processes are automatic, unconscious, they can account for emotional contagion, and all work with direct perception of emotional cues. By contrast, the slower processes act by inferring indirectly the affective state from cues, not requiring direct perception of cues of emotional experience in the target, and thus are thought to involve more advanced cognitive functions. Although its role in empathy is debated (Baird, Scheffer & Wilson 2011; Blair 2011; Decety 2010; Gallese et al. 2011), mirror neurons are thought to activate the associated emotional representations when observing the target's situation and behaviour (Gallese 2003; Gallese et al. 2004; Iacoboni 2009; Keysers & Gazzola 2009, Preston & de Waal 2002), which can help empathy to occur. The perception-action model (Preston & Waal 2002; Preston 2007) postulates that perception of the target's state automatically activates the representations associated to that state to prime the observer with the adequate response, unless inhibited. This process works on episodic memory, plus it is also compatible with mirror neurons. Vignemont and Singer (2006) explains that empathy is sensitive to situational factors such as the target's emotion (it is facilitated for anger, sadness, or happiness, but less for secondary emotions, such as jealousy), familiarity or affective link with the target, shared experience of the target's situation, age, personality, or gender. They describe empathy occurring with "shared affective neural networks, which are activated when we feel our own emotions, and when we observe others feeling emotions" (Vignemont & Singer 2006: 440) using contextual information about the emotional cues for the appraisal. The accuracy is correlated to the similarity between the target and the observer's *experiential repertoire,* helping to reach similar evaluations. What all these mechanisms have in common, is that they rely on (even though not always explicitly stated) shared representations, experiences, and processes linking behavioural cues to mental and emotional states. These states in turn help activate the corresponding representations about the target's state, resulting in empathy when both parties' states appear to match from the observer's perspective.

The reported roles of empathy indicate its place in the duet of affect and cognition, and what it is about cognitively speaking. Vignemont and Singer (2006) suggests an epistemological role and a social role. Epistemologically, empathy is proposed to be a faster and more accurate route to predict others' behaviours. It would offer a more precise and direct way to predict their actions, because the activation of shared emotional networks gives a better access to

[1] Direct association and mediated association differ from the way the observer learns about the target's situation. In direct association, the observer directly perceives the target's situation or expressions. In mediated association, the observer perceives the target's situation through words.

motivational and actions systems. Sharing the emotional state means that we share the emotional significance of events as well. It means we do not need to live by ourselves an event to ascribe an emotion to it. If we see someone experiencing the event, or if we are told about it, we can associate an emotion to that event without experiencing it first-hand. Socially, empathy might also enable us to learn about environmental properties from others' experience. (e.g., seeing someone burnt by an oven will generate a negative association to that oven, which leads to an avoidance of it, without having to experience the event ourselves). By its social role, they suggest that empathy is related to morality even though empathy is not necessary nor sufficient for prosocial behaviours to exist. However, empathy can contribute to instigate sympathy, which can induce helping behaviours. On the other hand, empathy can induce self-oriented and negative behaviours. For example, if empathy induces negative emotions such as distress, it can close the individual from pro-social behaviours. Thus, Vignemont and Singer (2006) concluded that empathy is related to social coherence and social communication facilitation due to the chameleon effect: imitating others' actions, gestures and postures help create affiliation and fondness between individuals. This is consistent with Bailey (2022), proposing that empathy is a way to achieve humane understanding, and endorse others' emotions. The observer, by comparing what she feels with what she believes the target feels, and by noting the degree of concord between the two, can then offer the target to be humanely understood. Being humanely understood is non-instrumentally valuable for the target. In that way, empathy affects the target's wellbeing, because we have a need to be emotionally comprehended, allowing for the recognition of one's emotions, and because the observer will be able to better satisfy the other's emotional needs. As a way to understand and possibly accept others' emotions, empathy is also evoked by Oatley (2016). In summary, empathy fulfils a crucial role in the acquisition of information about the way others' minds work to predict their behaviours and acquire emotional significance of environmental properties. Empathy in turn helps understand them better and potentially provide adequate plans of actions for support.

Of course, it presupposes that we can accurately read other's emotions, and that we share common experiences, (probably in the form of representations), and it is not clear yet why empathy can be blocked in certain situations, even though we understand what the target is experiencing. To exemplify this, let us take an example: a co-worker we had a good relationship with has just been laid off. We see that she is sad and feel sorry for her. It is a relatively normal case. But now let us say that co-worker was a competitor (in this case, due to the difficult times, it was either her or us which would have been laid-off), someone we felt aversion for, and thought was despicable. When this person has been laid off, despite seeing her sad expression, we may feel satisfaction and relief instead of feeling sad for her. Chances

of us attributing a matching state are seriously reduced, empathy will not occur. Intuitively, it seems like goals are involved in both cases in the form of goal convergence or conflict of interests. Empathy is facilitated when the target's goals appear congruent with those of the observer and blocked when goals are in opposition. We shall come back to that in the later sections 3 and 5 when discussing engagement.

The mechanisms underlying empathy appear to be present at every stage of the appraisal process. Vignemont and Singer (2006) stated that current evidence cannot tell apart whether empathy is executed in an early or late stage of appraisal. Thus, empathy can take a faster route for empathic response, and then be inhibited when factoring contextual assumptions in a later stage (late appraisal model), and/or take a slower route where emotional cues are always evaluated when contextual information is taken into account for the appraisal at an early stage (early appraisal model). This suggests that empathy might be closer to an iterative and continuous process, rather than a single event situated early or late in inferential processes. As Cunningham et al. (2007) proposed, there could be an appraisal model in which a stimulus is reprocessed through an evaluative cycle. "In this cycle, stimuli are interpreted and reinterpreted in light of an increasingly rich set of contextually meaningful representations" (Clore & Orthony 2008: 14, on Cunningham & Zelazo 2007). It indicates that an unconscious and automatic feedback loop occurs from the first affective reactions and progressively become refined evaluations, more reflective, in other words, they become emotions. Following that, we reappraise sensory inputs in light of the new affective state. Applied to empathy, it implies that there is an *empathic loop*, in which there is a constant 'attunement' between perception and appraisal processes, between the target's emotional cues and the observer's emotional responses. That type of process implies that we do not spontaneously empathise at the first sight in a one-off fashion, but we constantly reappraise our perception considering the previous appraisals. Indeed, we gradually construct our emotional response to fit the perceived cues, based on relevant representations from our memory and key emotional features of the target. Later, we will take insights from that model to build a relevance-theoretical description of empathy, in which perceptual information processing and feelings through emotion elicitations alternate and interplay continuously.

Before moving onto narrative empathy, we will mark the important points seen so far in this section that we will employ later to build our model in section 5. Empathy is considered as an outcome, resulting from a goal-oriented appraising process involving inferences using our past experiences. Empathic responses are not aimed at nor sought after compared to other non-matching outcomes in the process. A single mechanism elicits all kinds of appraisals, which

can be non-matching or matching from the observer's point of view. If the conditions are met, an empathic response occurs. So far, the conditions to elicit empathy are:

(1) An observer attends to the target's situations.

(2) The observer infers the target's mental state from cues (perceived or inferred) of the latter's situation.

(3) The observer appraises the target's situation.

(4) The observer estimates that her own mental state matches the one she attributed to the target, and that her mental state is caused by attending to the target's situation.

Producing an appraisal similar to the target's inferred one requires a shared pool of representations with the target, as well as convergent goals. Attribution of mental states (mind-reading) is a matter of cue-reading, exploiting our past experiences. Experiences associate a cue to a specific mental state, allowing to draw mentalistic inferences about people's situation. Finally, the mechanism underlying empathy is a continuous appraising process. It involves a feedback loop between inputs and affect formation, refining the appraisal over time (thus constantly moving between matching or non-matching on a continuum). In the following sections, we will keep all those points in mind, and assume that appraisal processes and inferential processes are describing the same phenomenon from different perspectives. This assumption will be helpful to explain empathy within relevance theory later.

3. Narrative Empathy

Bruner (1986) defines narratives as a mode of thinking to treat agents, intentions, and vicissitudes of the intentions. It is opposed to a paradigmatic mode concerning explanations of how mechanisms and physical processes work. Narratives seem to stem from conversations (Bortolussi & Dixon 2003; Oatley 2004) to exchange about events of emotional significance (Sperber 1996). A large portion of literature and fiction, except for lyric poetry, involves a narrative framework (Djikic et al. 2013). "Literary arts" wrote Nussbaum (1998), citing Aristotle (1970) "show us 'such as might happen in human life' (1970: 5). This kind of art can affect our ability to imagine what is it like to live the life of another person" (1998: 5). Indeed, fiction has been proposed to be a simulation of the social world allowing exploration of others' minds and social interactions (Oatley 1992, 1999a, 2016) and offering models of the social world (Mar & Oatley 2008). Through reading of fiction, we acquire experience helping us understanding others, and thus improve our empathy.

Feagin (2018) also explained empathy as a simulation of others' mental states, in which we adopt others' perspective by using our own mind "to model target's mental activities under certain conditions" (2018: 146). According to her, sharing the same emotion as the target is not enough, it must also be produced by similar processes. Coplan (2004) proposes that we do not have the same knowledge as the target, so we *bracket* for the time of the process our thoughts, beliefs, sensory inputs and replace it with the one we attribute the target to have. Then, we run the simulation of her mind without any parasite processes that, we suppose, might not play a causal role in the target's mental activities.

Literature and empathy are intimately linked. Oatley and Djikic (2013) reviews a variety of experimental studies trying to measure effects of literary reading and engagement on empathy on short, medium, and long-term, using fMRI-based or a Mind in the Eyes paradigm[2] showing an overall increase in empathy after reading literature. They hypothesised that fiction encourages us to explore and draw inferences as we construct mental models of complex characters (Oatley & Djikic 2018). Oatley and Djikic (2013) suggests that fiction offer a gentler and risk-free environment for practising empathy without experiencing the negative consequences that we may encounter in real-life, such as misunderstandings or upsets. Thus, fiction facilitates the development of social and empathic skills.

Narrative empathy is "*sharing of feeling and perspective-taking induced by reading, viewing, hearing or imagining narratives of another's situation and condition*" (Keen 2014: 1). In narrative empathy, as opposed to empathy, we can empathize more easily with a fictional character, as narratives might give us more direct insights on the characters' mind as opposed to real life (Keen 2010), and since we feel those emotions in empathic processes, it creates in the reader a sense of intimacy with the characters (Mar & Oatley 2008). Fictional characters, by definition, are not real. So, to explain why we feel emotions for them despite everything, it was advanced that witnessing situations or imagining emotional states in others is sufficient to activate automatic representations of the same state in the spectator (Preston & De Waal 2002). Primitive empathy was linked to other fast-and-dirty life-saving responses and suggested to provide a first quick feeling response to learning another person's emotional state before a more deliberate role-taking (Keen 2010). Another aspect potentially making empathy towards fictional characters easier is that fictionality renders social contract between the reader

[2] Mind in the Eyes test is a non-self-report test of mentalising elaborated by Baron-Cohen et al. (2001). It is composed of 36 photos of actors' eye regions to which participants must attribute an experience to the photographed person among four possible mental/emotional states. It measures the ability to make mentalistic inferences, and thus can be used to measure perspective-taking, an aspect of empathy.

and the characters impossible, and thus no social obligations or conflict of interest can impede on empathy.

Hogan (2003) presents two types of narrative empathy: situational empathy and categorical empathy. Situational empathy concerns empathic response caused by situations lived by the target similar to our past experiences, and the second concerns empathy based on perceived similar characteristics with our own. Categorical empathy was presented as more widespread whereas situational empathy as the more ethical, but less reliable kind because it requires the observer to have lived comparable experiences before. Those categories seem to overlap with Vignemont and Singer's (2006) account where we have many factors facilitating empathy such as familiarity, shared experiences with the situation, similar apparent characteristics. We could see all those types of empathy and factors pointing to the existence of clues allowing the transfer of our own experience to the other's situation, and even characteristic-related cues are comprised within this projection as they may be the more direct indicators of shared experiences.

Narratives elicit emotional responses just like real events, though there are a few differences in how both are processed. Konijn et al. (2009: 14) discusses that we use the same psychological systems for mediated/fictional events and factual/real events and that both systems generate similar neural responses. Fabb (2022) argues that although we use the same system for fictional and real events, we could still process fictional events differently from real ones. Indeed, knowledge and representations acquired in narratives interact less with factual representations. Thus, fictional representations are mainly connected to other fictional ones. Real events and people have greater personal relevance to the reader than fictional ones due to reality being more relevant to us personally and to our existing knowledge. In other words, fictions usually do not achieve external relevance, because we know they are not real. This also explains why our emotional engagement produces behaviours in fiction different from real events: we do not run away from the train seemingly coming onto us on screen, but we still observe behaviours to intense fictional events, such as closing our eyes/the book, leaving the theatre, or interrupting the story, ending the narrative transaction in cases of extreme distress and intense emotional reactions (Keen 2012). In this view, empathy works with fictional characters as well (even though they are not real and thus do not really feel anything), because we can map our own knowledge of mental lives to what we assume the characters know within the story, and from there assume what they could be experiencing, and bracket this experience, as suggested by Coplan (2014), as of the characters' from their point of view

rather than ours[3] . Ultimately in empathic processes, the emotional responses are ours, not the characters', and empathy requires that we remain aware that the emotion is due to witnessing their situation, as we suppress any *actual* behaviours as for example, jumping onto the scene (Gerrig 1993). Additionally, Fabb (2022) observes that during fictional engagement, we can experience certain negative arousals and put these responses within their context using metacognition (we realise that the responses are not due to *real inputs*), and still appraise the experience as positive because of this decoupling. Rickard (2004) proposes that music stimuli and real-world stimuli both provoke similar emotional responses and arousals, indicating an underlying functioning of emotion elicitation. In an analogous way, we can suppose a common mechanism being responsible for processing both *real* input and *fictional* ones, with the only difference of treatment being that people factor the factuality of an input when making inferences.

Narrative techniques are thought to trigger narrative empathy, but as Keen (2006) mentions, this parameter alone is inconclusive. Narrated monologue, psycho-narration, free indirect discourse, those are different way to represent inner life. The choice of the person also influences our empathic responses: first person seems more authentic, but it is not clear if it is truly more adequate than the third person to trigger empathic responses. Indeed, older readers might be less prone to empathise in the case of first person than younger readers, and more with the third person. The first person may allow greater empathy for unreal situations, by releasing the reader from self-protection through suspicion and scepticism, by looking more genuine. Empathic responses can also be provoked through narrative uses of formulaic conventions (e.g., thriller or romance) resonating with past narrative experiences or with use of unusual of striking effect encouraging foregrounding and paving the way to empathic responses (Koopman & Hakemulder 2015). Stylistic features are hardly the central reason why we feel narrative empathy. However, they do seem to contribute to the activation of narrative empathy in certain situations, through foregrounding and ultimately literary resonance, by promoting self-reflection (Koopman & Hakemulder 2015), or triggering surprise (Fabb 2022), thus putting the depicted events into perspective with our own experiences. For those reasons, we will not consider stylistic features for this analysis, though we must keep in mind that certain linguistic devices might encourage self-reflection and empathic responses.

In cognitive narratology, narrative emotions are classified depending on the way they are elicited. Mar et al. (2011) distinguish between *evoked* emotions and *character-driven* emotions.

[3] And put aside our additional knowledge we could have as readers which we assume the characters do not have, to better simulate her states.

In the first category, emotions are derived due to them being rooted in memory. Resonance with past experiences, themselves loaded with emotional content, leads to emotion elicitation. In the second category, emotions are elicited through the engagement with fictional characters, through appraisal of events or emotions through said characters. From there, three types of character-driven emotions were identified: emotions of identification, emotions of sympathy and emotions of empathy. Emotions of identification originate from when we want to be or feel like being a character. Identification is generally thought to invite empathy (Keen 2007) and is distinct from it (Keen 2012). Emotions of sympathy encapsulate any emotion we feel for a fictional character, but that does not match the one the character is having. Emotions of empathy come from when we somewhat feel the same as the target character. As we are interested in narrative empathy, we will focus on the last type of emotions. However, we need to keep in mind that those three types of emotions can overlap and interact with each other: identification can lead to more sympathy and/or more empathy for a character, more empathy can lead to more sympathy and/or identification and vice-versa. Character identification is a process that is often associated with perspective-taking, by offering the reader characters' circumstances, actions, thoughts or motives, speech (reported or inferred) as hanging points to transpose her own experiences and opening the door to potential empathic responses (Keen 2015).

Cognitive narratology explains narrative empathy as being part of a larger mechanism using models and experiences to infer others' emotions. Oatley (1999b) proposes the planning processor as the main mechanism involved in character-driven emotions and thus in narrative empathy. It is a cognitive mechanism allowing one to simulate the social world by selecting goals and form intentions from goals and make plans for our next actions, to best accomplish those goals. We can also *enter* other people in the processor, and map them to their goals, beliefs, and other relevant information to simulate what will be their behaviour and act in consequence. When reading a story, this same processor is used, but since the reader withdraws herself from the real world, she will suspend her own beliefs and goals. Instead, she will use the fictional characters' goals, beliefs, and actions with the planning processor to feel different kind of emotions for them as their goals are met or not. As the planning processor serves to run simulations of social interactions in our everyday life, it is not surprising to see that fiction's primary function was proposed to be acquisition of social experience through abstraction and simulation of social exchanges in the realm of narratives (Mar & Oatley 2008). As we use the same brain areas to comprehend people as to comprehend stories, this gives some credits to a mentalising mechanism to simulate complex social situations (Oatley & Djikic 2018). Indeed, in empathy, it was proposed that we understand people's motives though

a model of their mental life and feel something close to what the character feels, but we identify this emotion as our own and that of the character (Mar et al. 2011). Oatley (1992, 1999a) further proposes that narratives themselves are a simulation of mental states and motives, whose progression is evolving throughout events in the narrative. In other words, we use the same cognitive abilities to understand people in the world as to understand fictional characters, their plans, and their motives. The notions of mental models and mentalistic inferences are central to understand character-driven emotions as they are elicited through characters' mental modelling and appraisal of events and their relation to characters' goals, and certainly the reader's goals to some extent as well.

3.1. The role of engagement

Engagement in a story was highlighted as being crucial to narrative emotions, transportation (Oatley 2016; Koopman & Hakemulder 2015) and empathy (Coplan 2004). Transportation was proposed to be a state in which the reader of a narrative entertains vivid simulations of depicted events within the narrative (Gerrig 1993; Green & Brock 2002) involving more simulations of social experiences (Johnson 2012), various emotional responses, namely for characters. It is also sometimes called "absorption", "narrative engagement", "narrative emotions" or "involvement with characters" (Koopman & Hakemulder 2015: 90). Koopman and Hakemulder (2015) used the term *transportation* when the reader feels being pulled into a narrative world, whereas the term *narrative empathy* highlights character-driven emotions and the affective link the reader has for characters. In turn, transportation was linked with narrative empathy (ibid). The main factor allowing being emotionally transported or feeling emotions for a character is engagement (also called involvement, or investment). Mar et al. (2011) proposed that empathic involvement with a character slows down reading but enables more elaborate simulations of depicted events. However, Coplan (2004: 148) explained that in empathic engagement, even though readers can be deeply involved in the characters' experiences and can connect to them, they remain aware they are different from them, readers know they do not become their characters when empathising. The more we are engaged in a character, the more likely emotions may be elicited when witnessing her situation, and an empathic response is also probable as a result. Engagement intimately intertwines with transportation, inasmuch they could be different sides of the same coin. So, understanding what the inner workings of engagement will be crucial to understanding narrative empathy.

Oatley (2016) proposed engagement being the product of making mentalistic inferences and emotional transportation. As engagement in narratives was found to improve empathy and

social understanding, Oatley explained these effects in two accounts, one based on processes, the other one on contents. In total, Oatley proposes five components constituting engagement.

In terms of processes, Oatley (2016) suggested that two following processes are likely to be involved in engagement. The first one is inference-making about characters' feelings and mental states. He posited that we use "inferences of the same sort we make in conversation, about what people mean and what kinds of people they are" (2016: 621) and how they feel, and that by doing so, we try to attain "a deeper identification and understanding" (2016: 621). We already saw that fiction as intimately tied to the social world, and so, having engagement described as an activity involving inference-making about people's minds and feelings is coherent. Engagement as an inference-making activity might also indicate us how our engagement can vary among characters during a story. The second process is transportation, the degree to which the reader is immersed into the story, emotionally involved, and activate imagining the depicted events. The reader's attention is mainly focused on narrative elements at the expense of the surrounding world. During this state, the involvement in the story is such that narrative emotions are strongly felt and we experience the moment as if time flew faster. Thus, Fabb (2021) suggested that transportation could be a variant of Csikszentmihalyi's (1990) flow experience[4], presenting similar experiential characteristics. It was shown that emotional transportation is likely to have effects on the readers' empathy (Koopman & Hakemulder 2015; Johnson 2012). Bal & Veltkamp (2013) showed that emotional transportation in readers is related to an increase in empathy, but only after a certain time of incubation. It seems like the role of transportation in narrative enjoyment is also related to taking an emotional perspective of characters and sometimes feeling characters' emotions via empathy. Engagement is intrinsically tied to narrative emotions. The fact that we can be engaged in characters in different ways might also explain the different outcomes of character-driven emotions such as emotions of sympathy and emotions of empathy. It follows that the kind of emotions may depend on the type and intensity of the engagement we invest in a given character.

In terms of contents, Oatley suggested three aspects of narratives to explain the effects of engagement. The first one is literariness when it comes to character engagement. That is, the complexity of characters regarding their motives, emotions, and the exploration of their mental lives is proposed to provide the readers the impression of round characters. This is generally paired with stylistic devices to reflect on emotions such as foregrounding to defamiliarize the reader. Defamiliarization in turn produces aesthetic effects and emotions and favour reflection,

[4] The flow state involves intensive focusing on a challenging task using skill with immediate feedback from actions. It is described as an experience when one is completely engaged or absorbed into an activity, altering one's sense of time.

which participate to increase engagement. In other words, engagement can arise from wishing to experience complex emotions and vicarious mental situations. The second aspect is expertise about the subject of the story. The subject can vary across stories, most will generally emphasise on what people's intentions in social interactions. So, reading is a way to become more knowledgeable about other people's mental life, their goals, and maybe predict their behaviours. The third aspect is pluralism of experiences. That is, narratives can engage us into living situations, emotions through and for various people we would never have had the chance to encounter otherwise. In other words, engagement arises when one tries to 'put oneself into a character's shoes and experience different life scenarios, from which we, readers, can learn about how people's minds work. In the end, engagement can be seen as much as an active process (by making inferences to attribute mental states to characters) as a set of dispositions or goals we are interested to pursue such as to acquire knowledge (often about others' minds), entertainment, experiences beyond our reach and probably many more.

We will sketch a portrait engagement to show how engagement influence our appraisals of characters in a given situation, let us take a simple example: in the story of Peter Pan, we suddenly read about the protagonist Peter Pan being trapped in a giant spider web in a perilous situation. We will consider that we have a certain engagement in Peter Pan, though this engagement will vary across people for different reasons. Let us suppose that we have an overall positive engagement in Peter Pan due to personal preferences. In sum, we like Peter Pan as a character, we might identify with him to a certain extent, we care about him, we want him to survive and fare well throughout the narrative. We will consider engagement as a continual appraising process, linking a character's various information about his mental life, goals, and personality to various values and norms we have through mentalistic inferences. Seeing him in this delicate situation, uncertain about whether he will make it out of this, the danger slowly closing the distance to him, we might feel a narrative emotion close to a kind of fear for Peter Pan, because we care about him and his goals. However, we would most likely not feel the same if, let us say, the antagonist Captain Hook, was caught in the web. The audience will not necessarily feel fear, we might even observe positive emotions in them since Hook generally interferes with Peter Pan and poses a threat to him in general. Antagonists seem to induce, prototypically, overall negative engagement in them, by committing acts which we will generally appraise negatively, and by looking, prototypically, distinctly different from the target audience (blocking identification to a certain point). Since Hook is a malevolent adult, children might have more trouble empathising with him. Of course, there is a plethora of counterexamples in countless works of fiction showing that we can empathise with antagonists; we can even feel fear for Captain Hook. The gist is to understand why we

would appraise Hook distinctly from Peter Pan in the same situation and link it to the type or the intensity of the engagement we hold towards each. Engagement is thus what dictates how we will appraise characters' action in a given event. Conversely, an absence of engagement for certain characters will denote a difficulty to identify with them, to connect with them, and thus to feel emotions for them: we have troubles to relate them to anything we know or care about. For instance, if the restrained person was an unspecified, unknown, under-described commoner, with whom we cannot relate, the reader would probably react in a more neutral way.

The point is, we evaluate characters' situation differently despite them being in an equal position. We will consider that engagement is at the same time a set of personal dispositions (stemming from the weighted importance of their acts, and from stored appraisals associated to that character), and the way we will attribute mental states, goals, intentions to them. All this will influence our appraisal of their situation and will decide whether empathy arises or not. This view is compatible with Oatley's (2016), as engagement is seen both as a structure and a process.

Oatley's insights (2016) into the components of engagement, seen previously, indicate that engagement depends on what drives us to consume, enjoy, and get involved in stories. That is, we have a number of narrative goals when we come to process a story, which overlap with the five aspects of engagement we have seen previously in this section. When we enjoy stories, we might seek to better understand others (mentalistic inferences), to experience novel emotions and feel disconnected from our daily life (transportation), we may want to explore the life and events of interesting characters (literariness), to learn models about mental lives and the social world (expertise) and getting to feel like another person (pluralism).

For all those motives, I postulate that narrative engagement is intrinsically connected to narrative goals. A narrative goal is anything we expect to obtain from the appreciation of a story. Characters usually play an important role in the fulfilment of our narrative goals, and thus we can engage with them by projecting narrative goals onto them, and they might lead some of them to fruition, positively or negatively. We might have many more goals than those listed in the above, more personal, more precise, despite that, it represents a spectrum of the main reasons why we would engage in stories and find the activity of processing narrative relevant and worth putting efforts into it. When we engage in narratives, we put aside our daily goals and take those of the characters for the time being plus some narrative ones incentivising the efforts to process and make inferences to comprehend the story. This does not mean that we do not use our daily goals during narrative processing: characters often have issues similar

to ours. We can identify with them, and it allows us to reflect upon our daily goals, feelings, issues in a non-committal way through our narrative goals. The main issue is to conceptualise engagement in a cognitive way, as it is essential to transportation, character-driven emotions, and narrative empathy, but the portray sketched by cognitive narratology remains loosely defined and multifactorial. Nevertheless, *goals* are a well-known concept in the appraisal theories of emotions, and if we can define engagement in terms of goals and appraisals, then we might explain it from a relevance-theoretic perspective. To overcome this challenge, we will provide in section 5 a relevance-theoretical definition of emotional engagement (in characters or in a story) as a product of goal relevance in relation to the readers' goals.

4. Relevance Theory

In Sperber and Wilson's (1986/1995) description of relevance theory, the detailed utterance interpretation and information processing follow very simple rules and heuristics towards an optimised cognitive economy. Two main tenets, the cognitive principle and the communicative principle, are at the core of interpretative processes.

The cognitive principle states that comprehension processes minimise costs and increase effects in selecting contextual assumptions to maximise relevance. Relevance is a product of two heuristics: 1. Other things being equal, the less cognitive effort is required to process information, the more relevant it is. 2. The more positive cognitive effects is generated by some information, the more relevant it is. Cognitive effects are contextual implicatures produced by selecting a specific contextual assumption. General information processing is thus efficient using "fast and frugal" heuristics to economise energy and produce useful conclusions from perception.

The communicative principle states that ostensive acts create a presumption of relevance. To identify the intended import by the speaker uttering an ostensive act, we must first retrieve the communicative intention which will allow us to process the utterance expecting optimal relevance from it, greatly facilitating communication by boosting the anticipated gains from the comprehension process.

The information processing sequence must explain in which order assumptions are selected and when to stop looking for potentially more relevant ones. These are the two rules explaining how we achieve that (Sperber & Wilson 1986/1995):

a. Follow a path of least effort to gauge cognitive effects. Assess interpretations in order of accessibility.

b. Stop when the expectations of relevance are met.

Sperber and Wilson (1986/1995) argued that the speaker, rather than directly changing the receiver's beliefs, aim to modify the receiver's cognitive environment, and more specifically, make manifest or make more manifest a set of assumptions. Assumptions exist within the cognitive environment and can change to various degrees of manifestness as a result of the perceptual environment or events, or ostensive acts. A more manifest assumption has a higher chance to play a causal role or to be mentally represented. Let us compare cases of strong and weak communication from Wharton (2016: 23) to exemplify how they influence the manifestness of conveyed array of implicatures:

1. A: How′s work going?

   B: The boss is a bastard!

2. A: How′s work going?

   B: (Sighs wearily)

3. A′s colleague, who works on the next desk, catches A′s eye, sits back and sighs.

Throughout those three examples, we notice that they communicate approximately the same propositions (namely that B is unsatisfied or even angry with her boss) in an ostensive way, but each time more indeterminately than the previous case. In the first example, we easily parse a strongly implicated set of assumptions that are made manifest by B concerning her work conditions. In the second example, we can still approximately attain the same conclusions, given we possess background assumptions about B's working conditions. In any case, we will most likely deduce from her sigh that B is not satisfied of her work conditions, because the assumption of her weariness is strongly implicated, and therefore, it will be strongly manifest, while some others (for example, that she is dissatisfied with her boss) are weakly implicated, therefore weakly manifest. The third example is even vaguer in the sense it makes weakly manifest an even wider array of implicatures, and it is noticeably harder to retrieve the implicature that B is dissatisfied of her work conditions, and even more so about her discontent about her boss. Thus, in *example 2* and *3*, the construction of implicatures as in

example 1 will require additional inferences and knowing other information or have other manifest assumptions about B's work. We say that *example 3* "creates an impression rather than conveying a definite message" (Sperber & Wilson 2015: 23). *Impressions* are a change of manifestness in an array of propositions in the cognitive environment (Sperber & Wilson 2015). They generally generated in weak and indeterminate communication, making manifest a large array of weak implicatures. The typical example in Sperber and Wilson (2015) is *an ostensive sniffing before the sea scenery by the speaker to the receiver*, which is quite difficult to explain with words what is exactly communicated. They are descriptively ineffable and non-paraphrasable without loss of meaning because the propositions composing them are too numerous. They are *in fine* the main line of defence of a fully propositional relevance theory, as non-propositional effects are proposed to be, in fact, a wide array of implicatures, which are propositional at core. We will come back later onto non-propositionality, and different ways to tackle it. In a narrative context, all the above applies in very similar ways, except expressions are inferred (for example, if the story is written) rather than perceived, and every piece, word, utterance of the literary artwork are ostensive acts. Manifestness can explain why certain assumptions are more accessible and play a bigger role in inferences than others, and also why other assumptions can act below consciousness or even not act at all.

In sum, the relevance-theoretic model shows that information processing happens in a systematic and productive way, geared towards cognitive efficiency. Ostensive acts aimed at changing the receiver's cognitive environment, by acting on the manifestness of assumptions in the form of implicatures, on a continuum from strong to weak, and from one or just a few, to a wide array of implicatures. This will constitute the chassis of the model we will build upon in the next section. Thus far, we have seen how relevance theory treats inferences and communication, though it is still uncertain how emotional cues and situational factors are processed to elicit an emotion, that is how inferences can affect emotions and vice-versa. Yet, emotions and mental states are crucial to account for (narrative) empathy, and they do not seem to fit into the standard relevance-theoretic picture, other than with elusive impressions. We will see in the next section how what are different answers to the non-propositionality of affective states and include them into the model.

5. The model of narrative empathy in Relevance theory

In this section, we will develop a model which can explain narrative empathy using notions developed in the framework of relevance theory. Inspired by the appraisal processes and the

cognitive-narrative views on narrative empathy, we will build successive interpretational models starting from the relevance-theoretic model and improve it by integrating relevance-theoretic concepts until we arrive at a model capable of explaining the different outcomes of narrative empathy, non-matching response, and no affective response. The first iteration of the model will use procedural meaning and positive affective effects to create an affective-interpretational process. The second iteration will include metacognitive acquaintance as representations accessed by procedural meaning, feeding the first model to allow for mentalistic inferences. The third iteration will include goal relevance to be able to predict the various affective outcomes. We will call this model the Model of Affective Relevance (MAR). After that, we will compare the model to other psychological models of processing, such as representation matching and the perception-action model and show how their functioning overlap with the new model. After completing our model, narrative empathy will be described as a possible outcome of the inferential processing of a fictional character's affective situation, relying on procedural meaning to access relevant metacognitive acquaintance and representations, which allows the reader to infer the target's mental state from cues associated to it. In the final sub-section of this section, cases of empathy are explained in terms of goal congruence and assumed representation sharedness between the observer's and the target's goals. Narrative empathy works on the same principle, the difference being that the reader (the observer) possesses narrative goals fuelling the relevance of processing the character's (the target) situation and scaffolds a representation of their mental state through inferences.

5.1. Procedural meaning and affective effects

To account for various cases of non-propositionality, procedural meaning will be a key feature of our engine. The notion of procedural meaning was introduced by Blakemore (1987, 2002, 2011) to account for discourse connectors such as *but,* which are descriptively ineffable, not paraphrasable, and do not hold any truth conditional or conceptual content. According to her, those expressions achieve relevance by activating a procedure of interpretation to guide comprehension of the utterance, constraining the inferential processes in recovering a wide array of weak implicatures.

Wharton (2016) applied procedural meaning to expressives, such as interjections, and proposed that they encode emotional procedures. "Ouch!" does not mean "I am in pain". *Ouch* is non-propositional, is highly context-dependant, can communicate a wide range of implicatures, and it is ineffable as well. In that sense, they are akin to facial expressions and non-verbal cues, as they communicate emotional states, are usually cooccurring (we generally say *ouch* accompanied with a painful tone of voice and expression) and they are, to a certain

a degree, unconsciously and intuitively produced, and are non-propositional. Expressive meaning thus stands on a continuum between conceptual and procedural, and procedural meaning would be a much more widespread phenomenon. On the procedural end, we would find natural non-verbal behaviours. Halfway in between, we would find interjections such as *Ouch,* partly a natural signal and partly linguistic. On the more linguistic end, we would then find expressives such as *bastard*, which are ruled by linguistic conventions and bear some non-propositional content. This continuum also parallels Grice's (1957) continuum between showing and Meaning_{NN} (non-natural meaning) describing cases ranging from providing direct evidence to the intended import being gated behind the speaker's communicative intent: facial expressions would largely stand on the showing end whereas interjections stand somewhat closer to Meaning_{NN}, being parented with linguistic utterances. Wharton's underlying idea is that those expressions are natural codes bearing procedural meaning, which activates a module-specific emotion-reading procedure in the receiver's side to retrieve contents of the corresponding emotion. In other words, emotional communication is relevant because their procedural meaning helps us narrow down assumptions, by constraining interpretation according to the speaker's emotional state.

Emotional procedures from expressions function because emotions themselves are procedural. Crucially, emotions themselves can be considered as superordinate mental states gearing our cognition towards specific behaviours, looking for specific stimuli in the environment, and providing our cognitive environment with new sets of goals according to the situation (Cosmides & Tooby 2000, 2008). Wharton & Strey (2019) explained the mental events triggered by emotional states with the term *positive emotional effects*, linking feelings to cognition and show how both cooccur and complete each other. The exact same utterance in different emotional contexts will lead to very distinct implicatures, preparations to actions and sets of emotional responses. To take Wharton and Strey's example (2019: 262-263):

> Andrew and Grace have arrived in a small village in France, quite late on a summer's evening. They book into a small hotel which, apart from the owner, appears to be pretty much deserted. Since itis August, they assume this is because during that month most people are away from the village on holiday. Andrew asks Grace if she thinks they are the only guests staying the hotel. Grace replies:
>
> (19.13) I can hear footsteps in the room above us.
>
> Grace's response interacts with existing assumptions Andrew is entertaining to yield the contextual implication in (9.14):

(19.14) We are probably not the only guests staying in the hotel.

Let us change the context. Andrew and Grace live in a remote part of the countryside. There are no houses for miles around. The nearest building, some distance away, is a maximum- security prison. It is a stormy night and shortly before they go to bed they hear on the radio that a dangerous convict has escaped. In the middle of the night, Grace wakes Andrew and whispers (19.13) in a frightened tone of voice.

[…]

Given the new context, subconscious physiological changes over which they have little or no influence have put Andrew and Grace in a state of hyper- alertness.

Between those two scenarios, what changed is the context of enunciation of the same utterance, and with it, vastly different background assumptions yielding different conclusions. This contextual difference brought a noticeable change in the couple's emotional state, which in turn changed the way they uttered (19:13) with non-verbal cues, the way they interpreted the utterance, and influenced the manifestness of assumptions concerning their wellbeing and prepare them to act accordingly. Thus, emotions contribute to relevance by orienting interpretation into a specific direction, towards achieving certain emotion-related goals. It means emotions activate procedures acting on inferential processes to retrieve specific representations, which themselves can bear emotional content. De Saussure and Wharton (2020) consider emotions as evaluative devices as in appraisal theory, and they coined the term *positive affective effects.* They proposed that affective effects arise from descriptive ineffability and involve procedurality. These effects in turn activate experiential heuristics orienting cognition. Affective effects are described as "powerful boosters for the search for relevance insofar as they dramatically facilitate the identification of what is worth being attentive to" (de Saussure & Wharton 2020: 197) and also as "bridges between, on the one hand, feelings and sensations and, on the other, cognition as traditionally construed" (Wharton & Strey 2019: 263). We will keep using the term *affective effects* from now on, as it entails the notion of emotional effects. It is also more neutral and global as we can use it to describe the effects arising from affects, mood and other affective phenomena on top of emotions. In other words, emotions and affective effects impact relevance by boosting or directing attention towards optimally relevant cognitive effects. Non-verbal communication also exploits these processes, in encoding emotional procedures, treated by emotion-reading mechanisms, which can be also decisive in cases of empathy to infer what the target feels. Additionally, de Saussure (2021) suggested that in literature and stories, our emotional state is influenced by the artwork by evoking past experiences from our memory encapsulating specific emotions. A purely cognitive inferential process can thus instantiate emotions, and emotions guide those

inferences towards specific conclusions. Logically, empathy includes all the above procedural processes, as it is both inferential and exploits appraisals to initiate feelings, then it is used to better know what others feel. As seen in section 2, such interactions between interpretation of the situation and the resulting feelings constitute the base mechanism of what happens in the appraising process, and thus empathy.

Affective effects arising from emotions are not propositional, and the standard relevance-theoretic framework is limited in addressing those notions. Sperber and Wilson (2015) described a kind of mental processing resembling affective effects in terms of *patterns of activations* (2015: 138), which de Saussure and Wharton (2020) assimilated as procedures. De Saussure and Wharton described those patterns of activation as priming certain actions and directing our reasoning and relevance in specific directions. From their point of view, facial expressions cause a speaker and receiver to share feelings and sensations, which are primordial to cause specific cognitive effects. As we saw in section 2 above, appraisals and inferences appear in both affective and cognitive science; they are closely interconnected. It gets more interesting from this point on: de Saussure and Wharton proposed affective effects to go beyond the mere retrieval of cognitive effects. Indeed, they explained our understanding of the metaphor "Juliet is the sun" as being much more than retrieving conceptual content, it would require the receiver to have experiences of passionate love, or at least similar ones. The idea is that not all contents can be rendered as propositions, and thus, certain mental contents are intrinsically talked about in terms of sensations, or qualia. It ensues that any attempt to accurately paraphrase "Juliet is the sun" is impossible, not only because a wide array of weak implicatures is at play, but because the mobilised contents are fundamentally incompatible with propositions. If we know that Romeo passionately loves Juliet, and we possess prior similar experiences, our interpretation of the metaphor of Juliet as the sun depends on our intimate feelings associated with passionate love, such as the warmth or seeing the world around as illuminated. Their implicit idea is that procedural meaning entirely relies on *experiential reactions* (de Saussure & Wharton 2020: 18), stored in our memory, and are not propositional. Artworks, such as poetry, were also proposed to produce descriptively ineffable, unparaphrasable, non-propositional effects, called perceptual effects (Kolaiti 2019), involving conceptual and perceptual representations activated sub-attentively. De Saussure and Wharton (2020) also hinted that poetic artefacts could activate "'pure affective effects' that can be relevant in their own right […] without the cognitive effects" (2020: 202). Affects could constitute another possible output of inferences alongside cognitive effects, as a result from the activation of personal experiences; an insight shared by other perspectives, more on that in Section 6.

To support the integration of non-propositional effects, de Saussure and Wharton (2020) evoked a pre-conceptual dimension, in which expressives find a better niche: new-borns presented with a cotton-bud dipped in liquid sulphur held under their nose would produce an expression of disgust, an emotional reaction which does not depend on the concept FOUL (or similarly, a frog reaction to catch a fly entering its field of view would not rely on the activation of a concept FLY). What is maybe proposed here is a kind of association between what is automatically activated upon perception of specific cues and triggering an unconscious response. This will be crucial for our model, and we will come back on this topic later in the coming sub-sections and in section 6, about the perception-action model. Although de Saussure and Wharton (2020) mentioned that empathy, or some similar kind of simulation could not be the origin of emotional sharing because "[s]ensing someone else's emotional state does not happen through a scheme of inference, nor through logical steps of derivation." (2020: 200), we will nonetheless embrace the stance that emotional sharing is a type of empathy (both terms would be synonymous in certain cases), and that it does rely on inferences, though we mean inferences in a broader sense, which includes automatic, fast and frugal, unconscious network activations, and not only *formal* logical derivations and conscious reflective processes. As it stands right now, our model needs a new concept to account for the non-propositionality of experiences, especially about others' feelings and mental states.

At this stage, we present the current version of the model of affective relevance and the cases it can account for. However, we must keep in mind that despite that the following description can leave us thinking that affective and cognitive processes happen alternately, as in a loop model: such back and forth does not exist. We describe how one aspect affect the other separately to better highlight how they interact with each other, whereas both processes happen online simultaneously and continuously throughout an inferential event. Thus, affective and cognitive processes are synchronous and overlap in many instances. Initially, it starts with the perception of someone else's situation being parsed through relevance yielding processes, then inferential processes can elicit emotions through activation of past experiences tied to emotions. The same inferential "loop" starts over. Perception is again parsed by relevance (maybe new events have happened in the meanwhile), but this time, newly elicited feelings and emotions will influence the parsing and inferential processes in a new way. This in turn can activate previously inactivated memories. These new activations can modify the current affective state again. This divergence of affective state is a case of positive affective effect. This change of affective state will give rise to further positive affective effects and cognitive

effects in the coming inferential operations. It will also be influenced by new information from the perceptual environment.

What we just described so far could account for the interpretational process of any affective event (e.g., Interpreting "I hear some footsteps in the room above us" in a peaceful costal village vs in a gloomy isolated hotel) but it is not completely satisfactory in inferences involving perspective-taking (and thus empathy) where another character and other parameters, such as reader engagement or character's affective states, are considered. In the following sub-sections, the current model will integrate metacognitive acquaintance to incorporate characters' mental states as well as our relationships with them.

5.2. Metacognitive acquaintance and perspective-taking

In this section, we will present and elaborate the notion of *metacognitive acquaintance* to allow our model to make mentalistic inferences about people's feelings. The idea of this sub-section is to broaden the conception of metacognitive acquaintance starting from Sperber and Wilson's (2015) descriptions (see also Cave & Wilson 2018) and consider it with Wharton and de Saussure (2020) and de Saussure (2021) to include non-propositional representations involving attribution of mental experiences within this notion. I propose that these experiences are accessed via affective procedures during perspective-taking. The end goal is for our model to be able to make mentalistic inferences in the realm of relevance theory and provide a finer-grained description of emotional communication, thus moving further towards accounting for empathy. And if we manage to account for empathy, we can account for narrative empathy as we stated it functions virtually the same (cf. end of section 2, the four conditions for empathy).

Metacognitive acquaintance (MA) is one of Sperber & Wilson's (2015) three proposed ways to identify the speaker's intended import along with enumeration and description of a receiver. The intended import is an array of propositions that are to be identified by the receiver as the ones made more manifest to her by the speaker. The authors exemplify MA in the following way (ibid. 140):

> "For instance, as a result of the [speaker]'s behaviour, the [receiver] may experience a certain change in his cognitive environment, and identify this change, or part of it, as something the [speaker] intended to cause in him and to have him recognise as what she intended to communicate. In this case, what is needed to identify the array is neither enumeration nor description, but merely metacognitive acquaintance."

Sperber & Wilson (2015) also specify that MA is about the "psychological awareness of the effects of other minds on our own" (2015: 140), focusing on its intentional aspect. However,

later in the same paper, "Juliet is the sun" is said to be understood by MA as well, "by attributing to the communicator's intention what they mentally experience" (2015: 147), due to a change in cognitive environment as a result of a speaker's change in their behaviour, but the speaker does not necessarily need to have *intended* the receiver to draw that specific inference for MA to work, as explained "her intentions may concern only the general drift of the addressee's inferences and remain quite vague, and so may the addressee's understanding" (2015: 147).

In the case of literature, Cave & Wilson (2018) discuss how sensorimotor and echoic effects (which are non-propositional) resort to our own experiences to infer *what's it like* to go through this depicted situation, and it is attained through MA, aiming for cognitive alignment and not a "duplication of thoughts". De Saussure (2021) suggests that our affective reactions come from intimate psychological experiences, which could be the essence of MA. He describes sharing impressions as about sharing emotional experiences, leading to establishing a sense of MA about others' minds. So, MA is proposed to lead people to realise that other individuals share similar experiences. The gist is that metacognitive acquaintance stems from experiences and by sharing them, they allow us to understand others' mental states, including ours.

From that description, we can list the properties of MA and posit on its nature. To identify the intended import of a behaviour, there are two phases: first, a change in the receiver's cognitive environment, and second, the recognition of the changed array. This last part is MA. MA is involved in cases of weak communication, in the sense it communicates very weak implicatures. Intentionality is not one of MA's essential features, but it can be in many cases. MA is prompted by the speaker's change of behaviour, so there are two implications we can draw from that. First, it is perception-based since the change of bahaviour needs to be noticed. Second, by *behaviour*, we can encapsulate tones of voice, facial expressions, non-verbal gestures, which we saw as encoding emotional procedural meaning in the previous sub-section. *Behaviour* also involves other ostenstive acts and literary examples, such as "Juliet is the sun" or certain sylltistic figures encouraging us to draw further inferences within the context (Wilson 2018). It follows that MA requires cues to work. By cues, we mean it can be either a perceptual cue (as with non-verbal cues), or an inferred cue (as with literary excerpts). However, we can also argue that even perceptual cues require some degree of inferences, such as perceptual inferences. Conversely, we can argue even inferred cues contain some percepetual elements to them. Another property we can deduce is that the change in the receiver's cognitive environment must happen in an automatic way. So, only then this same change can be noticed by the receiver. And lastly, MA has an experiential component, meaning that it is acquired first before being used in processes. So far, it seems like MA is made possible through very low-level processes. Then those processes automatically induce a change of

manifestness in a wide array of implicatures in the cognitive environment from attending others' ostensive behavioural cues (and the change does not necessarily have to be intended by the speaker) without any conscious input from the receiver to process it. What is concerned by MA is based on past experiences of attributing mental states to people.

From all we have seen up to now, we postulate that metacognitive acquaintaince are the representations involving attribution of mental experiences to other people or oneself, based on specific experiential cues. There are two types of representations allowing metacognitive acquaintance to work, each one operating in two subsequent phases, in order: first, perceptual representations, and second, metacogntive representations. A perceptual representation is composed of a one-way inference relation between a perceptual cue and an experiential state. A metacognitive representation is composed of a one-way inference relation between an experiential cue and a mental state, such that, if a presented cue significantly matches that of the representation, we will automatically infer the corresponding mental experience (and thus mental state). A cue can be anything recognised as input from perception, inference, a feeling, or a change in the cognitive environment. A priori, perceptual representations are not metacognitive, because their cue is perceived of an inferred perception (a *smile* is not a mental state). So, it is not *about* a mental experience, whereas metacognitive representations are metacognitive, because their cue is linked to a mental experience. By experiential state, we mean that the output of perceptual representations is a general type of *experiential* effect. The kind of effect is decided by the type of mental experience that happens upon the creation of the representation. During the creation process, salient cues are associated with the ongoing mental experience. For example, if pleasant and happy moments happened when we attended to smiles in others or in ourselves, we will associate *smiles* with the *experience of happiness*.

Experiential effect can be any kind of effect that can influence cognition, as well as physiology, affects and actions. In that sense, experiential effects can be cognitive, as traditionally envisioned in relevance theory, but also affective, perceptual, sensorimotor, and maybe even other effects such as physiological, or motor effects. We will come back to the nature of experiential effects when we discuss the perception-action model in section 6. We use the term *experiential effects* in a similar way to *responses* in their model. *Response* refers to a general range of phenomena, influencing somatic and autonomic systems. Perceptual representations possess an experiential effect as an inferential output, whereas metacognitive ones possess an attribution of experience as output. An output can always serve as an input to activate another representation, given it has a cue of activation corresponding to the output. The functioning of those two kinds of representations is very similar in their principle, they only differ in terms of the type of input (the cue of activation) and the type of output. Both types of representation

are successively needed to identify mental states in others' behaviours and situation, and we continuously draw this kind of inference in a spontaneous and effortless fashion.

In an emotion-reading context, a behavioural cue acquiring salience in our cognitive environment will activate the experiential content associated to that cue, and this content will in turn constitute an experiential cue to activate metacognitive representations associated to that cue. Once activated, the metacognitive representations allow for inferring/attributing mental states to the speaker who uttered the behavioural cue. For example, a *standard smile* will automatically activate perceptual representations linking a *standard smile* to the affective experience of happiness related to the smile in the receiver's mind, hence the output *happiness* experience is an experiential cue. This experiential cue activates all metacognitive representations with this similar experience as a cue. Once metacognitive representations are activated, an attribution of mental state is realised. Here, in the absence of other contextual cues that could inhibit the following response, it will prototypically prompt an attribution of the speaker being happy. That identification and attribution process linked to experiences is metacognitive acquaintance.

In this view, perceptual representations are responsible for perceptual inferences. I propose those inferences are due to representations as in the perception-action model (Preston & de Waal 2002). Cues activate networks of representations in an automatic and unconscious fashion, and do not necessarily generate a *positive cognitive effect* in the standard sense: the output is an associated somatic or autonomous response, which includes affective response. Those cues and responses are essentially non-propositional and based on past experiences, but they still generate a conditional relation between an input and an output and induce a *change* in the cognitive environment. I will argue that the essence of this *change* (prompted by activation of mental experience or a response in perception-action model terms) is procedural meaning, though we will discuss that alongside the perception action model and perceptual representations, and what it means for relevance theory in section 6.

As we see, there are two phases in identifying a behaviour to attribute a mental state. MA, as already described, mainly concerns the second phase. To clarify correspondences, we describe the second phase (the one involving metacognitive acquaintance), in relevance-theoretic terms, as identifying the change in the cognitive environment as part of the speaker's import. This equivalent to say, in representational terms, that an activation of a set of mental experiences in turn activates the relevant metacognitive representations tied to that experience. Both descriptions consist of and result in attributing mental experiences, to others or oneself, from a shift in one's own mental states.

If we take the example of the *standard smile* again, there would be two kinds of representations successively involved in recognizing the import. The first kind to arise is the perceptual representation: the *standard smile* is a cue activating perceptual representations linking *standard smile* to an affective experience characteristic of the experience of happiness. We associate *smiling* to *feeling happy,* because we personally experienced smiling as pleasant in the past. The sole fact of smiling is enough to produce a pleasant sensation influencing our experiences through facial feedback (Soussignan 2002) facilitating the association between *smiling* and *feeling happy*. Activation of the experience of happiness would constitute an automatic change in the cognitive environment, and I suggest that in this very example, it constitutes an affective effect. This type of associative representation is corroborated by various "primitive, automatic, … and involuntary" (Hoffman 2000: 36) mechanisms, such as mimicry, classic conditioning, and direct association. In mimicry for example, the observer automatically imitates the target's emotional expressions or non-verbal cues. By afferent feedback, the imitation causes the observer to feel the emotional state associated to that expression. In classic conditioning, if you were bitten by a god and got scared in your childhood, made you scared, finding yourself in a similar situation later, in other words, presented with similar cues, you will feel the emotions associated to that event (in this case, fear). Direct association means that seeing a situation or the target's emotional expression will activate past experiences associated with that emotion. The discovery of mirror neurons also supports this view. Observing the target's situation and cues will automatically discharge the neurons responsible for first-hand experience of that emotion (Wondra & Ellsworth 2015). As such, perceptual representations are also responsible for emotional contagion in our view. Note that emotional contagion is *not* empathy according to our definition of empathy, as empathy requires a metacognitive attribution on top of feeling a mental state. However, emotional contagion is necessary to produce empathic outcomes (Wondra & Ellsworth 2015; Coplan 2004). All those mechanisms are not *metacognitive* per se: although they involve feeling oneself, they do not involve attribution of mental states. However, those mechanisms are required to trigger the automatic *change* in our cognitive environment necessary for mental states attribution.

The second kind of representation is the metacognitive representation. If the experience of *feeling happy* is activated in our cognitive environment, this feeling will constitute a cue to activate all the metacognitive acquaintance possessing *happiness* as a cue. Then, the metacognitive acquaintance activated that way triggers inferences about what the target mentally experiences when performing a *standard smile*. We can attribute to the other a mental state different from the one we experience due to the change in our cognitive environment. The attribution may differ, but it can also be the same: we can experience a change of mental

state that matches our attribution of mental states, and we would know that this change is due to witnessing someone else's situation. This is extremely similar to how Vignemont and Singer (2006) articulates empathy in a narrower sense, as it requires both to feel and to know the other is the source of one's own state (2006: 435). It is also similar to Coplan (2004) who proposes that empathy must include a clear self-other differentiation, and in that sense, it overlaps and goes beyond emotional contagion by including simulative processes and perspective-taking of others' mental states. MA, I argue, is the fuel of mind-reading and theory of mind abilities necessary for perspective-taking, since it allows to infer mental states in others.

I suggest that there are different ways through which MA can be acquired, all varying in the degree of directness in living an experience, forming a continuum between firsthand to secondhand experiences. Some of these representations can relate to direct experience of a situation we once lived, akin to situational empathy (e.g. Seeing someone getting burned by touching a stove will elicit a more impactful empathic response if we already personally lived that event before), or restricted to only specific people, based on direct witnessing of their response to certain events (e.g. After seeing my friend displaying a stank face of disgust after eating almonds, I know that next time she will eat almonds, she will probably have a bad experience) or to secondhand information of their responses to certain events (e.g. My friend's companion tells me that my friend hates almond). It is also consistent with Mar and Oatley (2008), arguing that fiction offers us models and simulations of the social environment, through simplified, abstracted and compressed representations. They argued that this form of learning through fiction helps us acquire gives us specific social information through narrative experiences, to better understand other people's minds and can increase our capacity for empathy and social inferences. Thus, narratives help us link certain cues to what people experience. So, I argue that the kind of knowledge acquired through fiction most of the time is MA, due to its identical role in perspective-taking and inferring others' mental states. On top of attribution of emotions, affects and mental states, MA more broadly includes people's goals, intentions, desires, knowledge, and to personality traits, as they are mentalistic in nature and necessary to gain insights into others' behaviors and mental states in a given situation.[5] In the optic of a planning processor, MA is the fuel for mentalistic inferences, perspective-taking, and consequently empathy. However, there is a considerable difference between inferring a

[5] A consequence of intention being tied to metacognitive acquaintance is that cases of overt communication (Sperber & Wilson 2015) relying on intention recognition must use metacognitive acquaintance to infer that the speaker intends to communicate something (communicative intention) based on cues or behaviours provided by her. The attribution of intention to cues may vary across contexts and individuals.

mental state and feeling a mental state as a result of inferences. So, what distinguishes perspective-taking and empathy is, as we explain later, a matter of goals.

In this second version of our model, we integrated the notion of metacognitive acquaintance, relying on perceptual and metacognitive representations. This add-on allows us to do two things: one, to clarify the nature of procedurality, and two, to account for mentalistic inferences and perspective-taking. When cues are made available to our cognitive environment, they will activate an array of representations with the corresponding cue of activation. Following that, those representations (perceptual representations at this stage) will generate an experiential effect acting on physiology or cognition, that can be used in turn to activate other representations with an experiential state as a cue and which attribute a mental state (metacognitive representations). Procedural meaning here is equated with the mechanism of cue-activation, explaining that the large array of weak implicatures we usually observe in non-verbal communication is due to an activation of a large array of representations sharing a similar cue. So far, we obtained interesting insights into how automatic processes make emotion-reading and attribution of mental states possible, and appraisals might also heavily rely on the presented mechanisms to elicit emotions automatically and subconsciously. Thus, it makes it possible for an observer to spontaneously feel an emotion from seeing a target, and attribute that the target is feeling the same as she is. It fills the criteria to qualify for empathy. However, we still cannot predict when an empathic response should happen with certitude, as we do not know what makes the observer think that her appraisal converges enough with that of the target to attribute a matching mental state. We can have the vague intuition that if the observer assumes that she shares similar experiences (i.e., similar representations) with the target, it will increase the chances of her coming to an appraisal similar enough in appearance to classify it as empathy, but we cannot explain beyond that yet. In other words, we cannot explain why we can empathise with a colleague in some situations and not in others, despite having similar experiences. We will introduce goals via *goal relevance* into our model to untangle cases in which the observer attributes matching or non-matching states.

5.3. Engagement as product of goal relevance

In showing that the notion of *relevance* in affective sciences and in pragmatics were closer, Wharton et al. (2021) suggest that Wilson and Sperber's (1986, 1995) two principles of relevance were only two epistemic goals among a larger array of goals, where affective goals would also figure, called *goal relevance*. Any event is appraised in relation to its influence, positive or negative, on our current goals. Pragmatic principles of relevance can be equated to goals, such as of energy conservation to minimize efforts. Thus, they showed pragmatic goals

being able to prompt appraisals to elicit emotions in the same way as affective goals in certain cases (Wharton et al. 2021: 265):

> Imagine for instance that, on a Friday evening after a long week of work, you finally arrive home and are more than ready to open a beer and start resting. Suddenly, you realize that it is actually the last day to return your tax files, or that you forgot that you need to revise a paper this evening, or do some other demanding task requiring an unexpected effort. At this point, you may undergo a negative affect such as weariness. We find it plausible that this is an affective reaction to a stimulus appraised as obstructive to the goal of minimizing energy.

Since goals are essential to appraising events, we will show in this sub-section how goal relevance is essential to engaging with characters, involving character-driven inferences using MA, assessments of events according to our goals, and comparing our appraisal with the one we assume the target is experiencing.

I propose the following: our emotional engagement for a character is the product of the assessment of relevance of her goals, actions, beliefs, characteristics, relative to our own goals. By *our own goals*, it can be as much narrative goals as other non-narrative goals, related to our values, desires and resonances with personal experiences more generally. Goals will be used as an umbrella term to include "concerns, urges, plans, ideals, what is desired, needed, or is the object of any other motivating mental state" (Wharton et al. 2021: 265). Narrative goals encompass any motive that the reader might have when providing efforts to attend a story. Most commonly, we follow a narrative for enjoyment and appreciation (Mar et al. 2011), or Oatley's (2016) five components of engagement, but this ranges much further as we can follow stories for uniquely personal reasons, such as to better understand an aspect of ourselves, to discover a classic story, to be 'transported' into a different world, or because we know that one beloved character is featured in. In narrative goals, we can also include the reader's desires, concerns, and preferences of outcome about the unfolding of the narrative, which "often counters to the concerns and preferences of the characters, including protagonists" (Coplan 2004: 147). Scherer (2004) had a similar proposal for distinguishing between utilitarian and aesthetic emotions, suggesting that the latter involves appraisal less in relation to goal relevance and more to cultural norms and personal values. Under our view on engagement, norms and personal values are all encompassed by goal relevance, the difference being that narrative and aesthetic emotions involve another type of goals: narrative goals[6]. The idea is that readers continuously monitor the goal relevance of events in a story in relation to their

[6] Alternatively, the notion of "aesthetic goals" could also exist and be better fit to account for appreciation of other art media.

own and the characters' goals. The reader's emotional engagement in fictional characters will constantly evolve according to her appraisal of their actions and their contextual relevance regarding the progression of the story, the characters' goals, and the reader's goals.

Following that, let us explain why in certain cases we feel emotions when the progression of a character towards her goal (that is not ours) is affected, feel emotions for others but not the same as theirs, or sometimes no peculiar emotions at all, or sometimes empathise with certain characters. The idea will be that character-driven emotion elicitation is mediated by the amalgamate of all goal relevance assessments of character information. This amalgam is the engagement we hold for a character. In this case, we talk about engagement as a mental structure, containing all past appraisals for that specific character. Engagement stands on a continuum of positive (e.g., for protagonists prototypically) or negative (e.g., for antagonists), but we will see that it can be more complex. If we take the example of Peter Pan again, we can posit that, through identification, the reader projects some of her goals into the character of Peter Pan. By reader's goals, it can be her narrative goals (e.g., to be entertained, to better understand the social world, to see an ideological debate through narration, etc.), or more general affective ones linked to our values, ideas, and desires, and not directly linked to narratives.

We might project narrative goals onto Peter Pan, the protagonist, and we might also identify with him to some extent (e.g., because he is a power-fantasized representation of childhood, defeating evil adults, being brave, being able to fly. Ideas which children might identify with and appreciate; and for adults, because it reminds them of their childhood, the carefree attitude, the refusal to grow, or any other reason why Peter Pan would come across as a relatable character). The idea is that Peter Pan's goals are also ours for the time being if their fulfilment seems to coincide with the non-narrative goals of the readers (that is, goals outside the realm of the story or the reading situation). That is, we need a point of anchor, some degree of familiarity or similarity associated with traits of the characters to allow for identification. Finding shared traits should normally create positive affective engagement. In some way, we want Peter Pan to succeed in his undertakings. Some of the enjoyment of the story (i.e., the fulfilment of our narrative goals) will come from Peter Pan achieving his affective goals. Thus, characters function as proxies to achieve some of our narrative goals. However, we will see that narrative fulfilment does not come only from a goal correspondence with the protagonists.

By contrast, Captain Hook appears as a contemptible figure, getting in the way of the protagonist's goals. So, Hook is negatively affecting the progression of Peter Pan's goals, but not necessarily our narrative goals as readers: we will probably appraise him negatively in

terms of his moral acts and values (his actions and goals are in opposition to the reader's non-narrative goals), but it turns out we can still have positive engagement for him narratively speaking because he may fill his purpose as a villain, as a narrative character. We can be engaged in an antagonist as a fictional character for the quality of the writing, their entertainability, or the complexity of their personality, but less so for their moral stands. While we can be mostly "negatively"-narratively engaged in a character, we can have both types of engagement (that is, complex engagement) in one character depending on the situation and the goals at play. It follows that in the case of an antihero or a complex antagonist/protagonist/character, it is possible to cheer her in certain situations and jeer her in others, or sometimes both at once, depending on which story goal is at hand and how the character acts upon them.

To link this analysis with metacognitive acquaintance, we could explain engagement by saying that we are *prototypically* more likely to empathise with a protagonist than an antagonist because we are *generally* presented with more cues attributing goals to the protagonist which seem to coincide with ours. Conversely, the antagonist is presented with goals different or incongruent with the characters, and thus ours. To put it in simple terms, we tend to empathise with people to whom we can relate, who have seemingly the same goals as ours and positively contribute to their fulfilment. Conversely, we are less likely to empathise with people to whom we cannot relate, who seem to have different or opposing goals to ours. The narrative gap between readers' goals and characters' goals means that readers will only empathise when both *seem* to converge.

In a way, this echoes Carroll (2001) that we do not imagine ourselves being Peter Pan, we feel emotions for him and what is happening to him. We can feel strong emotions for them, but our emotions are not identical to theirs. It is because there is an asymmetry between us and the characters on various levels. We do not have the same information (readers often have more information than the characters) nor the same desires, we can develop preferences about the outcomes different from the characters. Whenever our readers' preferences are in opposition with the characters', it is impossible for us to match their mental state. To take Coplan's (2004: 147) example, the readers may not want an unrealistic happy ending, due to narrative considerations or information unknown to the characters. Thus, empathy will not occur with goals opposing to ours. Though readers have knowledge and desires different from characters, leading to dissimilar reactions, Coplan (2004) maintains that we can bracket our own agendas and knowledge when engaging with characters to simulate their mental state. So, we can represent their desires while possessing distinct desires ourselves in an empathic process. I propose that our own desires, preferences, and motives, integrated under the term *goals*, fuel the undertaking of the simulative processes of others' goals and mental states. If their goals

and simulated mental states seem to align with our goals and appraisals, it increases the chances of attributing a convergent mental state, leading to an empathic response. If they seem to run counter, it increases the chances of attributing a divergent mental state, leading to a non-matching response.

Engagement, as a structure and a process, entertains a very intimate relationship with metacognitive acquaintance. Indeed, MA is responsible for attributing goals and mental states to characters, which is needed to appraise them and make mentalistic inferences about them. Furthermore, metacognitive and perceptual representations as described in the previous sub-section also possess a "double form" like engagement. Representations are stored (i.e., structure form), and when activated, they produce an inference (i.e., process form). Engagement as a process can thus help us acquire metacognitive and perceptual representations about people, they help us associate situations and cues with character's mental states. Engagement as a structure is the crystallised network of those representations, whose parts can be selectively activated in regard to relevance, to understanding people's experiences. In sum, we can entirely replace engagement in our model with the representations involved in metacognitive acquaintance, as they virtually execute the same function.

To summarize, engagement depends on the readers' affective goals. In a narrative context, they can be of two types: narrative goals and non-narrative goals[7]. I argued that character engagement varies in terms of these two types of goals, that it is goal sensitive. In appraisal theory's terms, our appraisal of their actions or traits is a function of how they affect our goals in a given situation, triggering corresponding emotional responses. It can also be that the antagonist might fulfil our narrative goals just as effectively as the protagonist (e.g., entertaining us or fulfilling a power fantasy) but only differing in how they fulfill non-narrative goals. The reader may be engaged in an antagonist (i.e., she appreciates the character) despite not necessarily identifying with them. Non-narrative goals involving characters relies on factors such as similarity, familiarity, and relatedness interacting with the reader's goals. allowing identification with the character. Characters in stories are appraised by how goal-relevant their actions are to the reader's goals, narrative and non-narrative, and that is the key to constructing engagement. Seen that way, engagement is a goal-driven network of appraised events, establishing what we know about, what we feel and could feel of the object of engagement. Engagement could thus be applied outside the realm of narratives, the same tenets

[7] Therefore, it would make sense that there is a narrative engagement and an affective engagement, the first one being restricted to fictional characters, and the latter applying to both fictional characters and real people.

work with real people, except that we do not have narrative goals in mind when attending a real person's situation, only affective ones.

This is the final version of the MAR and we added engagement to distinguish different cases of affective state matching, including empathy. The initial chassis can handle general information processing but does not optimally support affective states. The first version can handle emotional situation and emotional communication with the add-ons of procedural meaning and affective effects, though it could not yet explain our ability to spontaneously draw mentalistic inferences and thus appraise people's situations. The second version can handle perspective-taking with the add-on of the metacognitive acquaintance. MA was newly redefined as a mechanism involving perceptual representations producing a change in the cognitive environment from a perceptual cue, and metacognitive representations automatically inferring a mental state from the said change, acting as a mental cue. Metacognitive representations are responsible for attributing states to people. However, MA without accounting for goals cannot explain people's appraisals of others' situation. The third and last version of the MAR can now handle empathy and narrative empathy after redefining engagement in terms of goal-relevant appraisals and as a network of representations prompting MA. In our model, empathy relies on inferential processes as in relevance theory, and automatic mentalising based on non-propositional inferences. An empathic response depends on our engagement in the target: the attribution of seemingly similar goals to the target, and attribution of a similar affective state to the target will ensure the impression of a matching appraisal, thus empathy. We tacitly equated inferential processes to appraisal processes and used affective effects, metacognitive acquaintance, and engagement (through goal relevance) to transition from purely cognitive processes to an intertwined picture of affects and inferences working as an inseparable duet.

Some aspects of the MAR may seem enigmatic, such as how appraisals happen in representational networks, how the representations accord with relevance theory or how perceptual representations work. To further illustrate those points, we will present other models from various backgrounds and examine how they can fill the gaps in the MAR and how they support it in the next section.

6. Discussion with other theoretic models

6.1. Representation Matching

Fabb's (2021) matching of representations paradigm enlightens us about the way we retrieve relevant contents in the empathic process to match perceived/inferred emotions. He proposed that our mind is constantly comparing internal schemas with perception. A schema "(plural schemata) is what we know, in general terms about objects and events" (Fabb 2022: 12) used to "predict what we will perceive, given an expectation that the world will generally be roughly as we expect it to be" (Fabb 2022: 13). There can be schemata "for objects and for events and for sequences of events (sometimes the term 'script' is used for schematic sequences)" (Ibid) and also "for our own mental activity, and for other people's mental activity" (Ibid). Fabb (2021) also proposed that when the perception is overly discrepant with our schemas, surprise is produced, and it potentially elicits an experience of ineffable significance. This happens with particularly discrepant objects, such as, mountains, because their large size exceeds that in most of our schemas. This can also happen with so-called perfect objects, that is, when perception corresponds exactly to the schemas. Such an event is surprising because we generally expect objects to mildly depart from the schema, and thus perfect objects are discrepant with our expectations of the way the world should work. Fabb described some of those experiences when encountering perfect objects as a fusion of nature with the mind, as they perfectly match. This closely resembles a form of aesthetics' empathy described as a "projective fusing with an object, which may be an object which may be another person or an animal but may also be a fictional character made of words, or in some accounts, inanimate things such as landscapes, artworks, or geological features" (Keen 2006: 213).

Fabb himself explained that narratives can be a source of strong experiences through empathising with characters. According to his account, empathising can produce both discrepancies and perfect matchings, the latter case giving an impression of fusing with the character. This coincides with Miall and Kuiken's (1999) suggestion of empathy as a "gap filling mechanism by which a reader supplement given character traits with a fuller resonant character portrait" (1999: 74) and empathy was even compared to metaphor (Douglas & Kuiken 2017 cited in Fabb 2022: 101), as there is a gap between the self and the other with an apparent bridge between the two. This suggests that empathy is the result of a representation matching device. Those representations can involve feelings, emotional content, sensations linked to past experiences. These representations about affective content, as I proposed earlier, are perceptual and metacognitive representations. The matching follows the standard inferential process of relevance to save effort and increase effects. The process is also helped by procedural meaning to narrow down the scope of implicatures, pointing to relevant MA. Implicatures retrieved through this empathic process are primarily non-propositional, since MA mainly concerns perceptual and mentally experiential content. So, empathy seems to

follow the same principle of matching representations to fit perception. Representation matching may be equivalent to the perceptual cues activating representations with similar cues of activation in the MAR. Then, we attribute whether what we feel corresponds to what we perceive (giving the impression that we *match*) or not. The way perceptions are matched and attributed to affective states will vary across individuals, because different experiences lead to a different attribution of mental states. It follows that everyone will come to different interpretations and affective reactions to events in the same story.

6.2. The Simulation Theory

The simulation theory and the model of affective relevance shares some interesting parallels. The simulation theory posits that our reasoning relies on mental models that we construct to understand the world (Craik 1943). Mental models need to be similar to the real-world occurrences they represent. This means that mental models need to be iconic, that they share, in a way or another, some structural properties with what they represent (Johnson-Laird 1983). This theory opposes the "language of thought" theory, in which logic and syntactically structured strings of symbols would be the basis of our reasoning, involving formal inferences using propositions (Fodor 1975; Pylyshyn 2003). Below is a problem to solve from Johnson-Laird and Oatley (2021: 3) illustrating how "language of thought" is refuted by specific spatial inferences.

− The circle is on the left of the rectangle.

− The square is below the circle.

− Therefore, the square is diagonally below and to the left of the rectangle.

If its premises are represented as strings of symbols, such as those from proofs in logic:

− On-left-of (circle rectangle).

− Below (square circle).

they call for complex rules of inference, or axioms, that make possible conclusions such as the one above.

Their answer is that mental models offer a significantly simpler way to solve that problem than logical derivations of propositions by mentally representing the spatial layout. Models come in different types with various degrees of abstractedness. Many of those models might be involved in a single simulation, such as a when people imagine rotating objects in their mental space (Shepard & Metzler 1974). This process not only involves spatial models but also kinetic models to visualise movements unfolding in time and to comprehend temporal sequences of events. Likewise, models may represent abstract entities, such as intentions and ownership which cannot be visualised (Johnson-Laird 1983: 416) but can still draw inferences from their structural properties. Those inferences about abstract properties may need increasingly difficult deductions requiring more models of different kinds for each inference (Johnson-Laird & Byrne, 1991). Oatley and Johnson-Laird (2021) also apply the simulation theory to explain how emotions can arise from processing abstract art by linking iconic models with associated emotional responses to aesthetic cues. Models about oneself are needed to assess the impact of an artwork on oneself, allowing access to explicit knowledge and then producing an evaluation of the artwork. This evaluation in turn can thus elicit an aesthetic emotion. Those iconic models are constructed from the perceptual processing of visual cues. Oatley and Johnson-Laird (2021) suggest that cues evoke emotions on the basis of mimesis, meaning that they mimic features of the way emotions are expressed or the emotional behaviours associated with the objects or events. Abstract art emotions can be understood in terms of perceptual cues activating associative models that can provoke an initial emotional change in the audience. The audience can then assess this emotional change using mental models about themselves. By the way, this emotional change greatly resembles the functioning of metacognitive acquaintance as well as our introspective abilities to produce more complex emotions about the artwork.

The simulation theory was proposed to be part of the functioning of empathy (Carruthers & Smith 1996; Davies & Stone 1995a, 1995b) supported by the discovery of mirror neurons, suggesting the possibility of actions and mental states coded at a neural level. Narrative (and fiction) processing was suggested to involve such simulations, and thus comparable models, to empathise with characters (Mar et al. 2011; Mar & Oatley 2008; Oatley 2016). Oatley (1992, 1999a) argues that fiction is not a description, but a kind of simulation. According to him, simulations are used to understand complex multifactorial situations where different processes are involved. Narratives indeed usually contain varying amounts of characters with their own motives and stakes, and all sorts of events may happen to those characters (Oatley & Djikic 2018). Following that line of thought, narratives require simulations to be processed. In stories, simulations of people's circumstances are achieved by focusing on the characters' concerns

and plans while putting aside our own. This allows us to imagine what it would feel like to be in a certain situation that leads to experiences and emotions, given that we are engaged in the stories and characters' (Oatley 2016). Emotions are "fundamentally important for stories." (Oatley and Djikic 2018: 5). This may explain why writers and authors thrive for accuracy to describe the world, so that the reader's simulations and the resulting emotions are accurate. By means of narratives, we can enter others' minds as they offer us mental models necessary to understand conflicts, people, and the social world. The importance of mental models could be reflected in the larger size of our brain, and our increased sociality compared to other animals. This pressure on individuals requires of them to possess a larger database to store mental models of interactions with allies, peers, and competitors (Dunbar 2003). Literary texts show us various social events such as incidents about characters' lives, their emotions and their interrogations. Out of them, we construct models through abstraction and simplification, extracting the gist in the form of straightforward models ready for use in future mental inferences (Mar & Oatley 2008). Mar and Oatley further propose that fiction improves empathy and social inferences abilities by providing models of others' mental states responsible for reconstructing others' affective states.

The MAR complements the picture of simulation theory in explaining narrative empathy. We took inspiration from Mar, Oatley, and Djikic (Mar & Oatley 2008; Oatley 1992, 1999a, 2016; Oatley & Djikic 2018), so similarities with the simulation theory and their accounts of models can be observed. Models, while not directly designated as representational, are nonetheless representations in their iconic structure and their identical role in initiating spontaneous inferences. Metacognitive and perceptual representations can be seen as equivalent to mental models in narratives, as they function virtually in the same way in drawing mentalistic inferences about others' minds, reflecting upon our own mental states, and allowing empathy to occur, as long as the observer assumes a pool of shared representations between the target and the observer.

An interesting result of equating models with representations is that it means that an array of representations of similar kind being activated at once can be a case of simulation. Following that, mentalistic inferences involving metacognitive representations can be considered as simulations of mental states. The MAR shows how models in the simulation theory are assembled as the relevance-theoretic framework can accurately explain how representations are accessed and how various goals play a role in predicting empathic responses.

6.3. The Perception-Action model

The MAR is also consistent with the account of the Perception-Action Model (PAM), and its account on empathy (Preston & de Waal 2002). It states that "the perception of a [target]'s state activates the [observer]'s corresponding representations, which in turn activate somatic and autonomic responses" (Preston & de Waal 2002: 4). Reponses "refer to large class of phenomena" (Preston & de Waal 2002: 4), including emotional, physiological, behavioural responses and other inputs for other cognitive processes. If we try to transpose those responses into relevance theory, they do not fit the idea of a cognitive effect understood in a propositional way,[8] but they better mesh with affective (de Saussure & Wharton 2021), sensorimotor (Cave & Wilson 2018) and perceptual effects (Kolaiti 2019) which are non-propositional. Initially designed to explain basic behaviours in animals, the model was also applied to social behaviours, and to empathy and its development in humans, showing a phylogenetic continuum from emotional contagion and altruistic actions to empathy.

The PAM predicts effects of familiarity, empathy-related disorders, and evolution of empathy throughout the development of children. This view indicates empathy stemming from a more general organisation of the nervous system, which "adaptively generates responses from perception, using the same representations to code objects and their associated actions" (Preston & de Waal 2002: 6) facilitating adequate behaviour to a situation. Then, the same organization was also used for social animals with some improvements to respond either with a matching response or an instrumental one depending on the situation. Representations are acquired and change with experiences, to the extent that they could be considered the crystallised form of experiences. A shared pool of representations among individuals explains effects of similarity and familiarity. The PAM thus predicts why individuals act and react a certain way, by helping or punishing a behaviour, due to the activation of an association between perception of a stimulus and a response. They proposed aversive signals, such as distress in humans and primates, to have evolved so that others will feel their distress as well and terminate those signals by aiding the distressed individuals. The others will act that way, because perceiving a distressed conspecific produces an unpleasant response, prompting them to act to cease the distressed cues. Thus, this basic system allowed for affective resonance to appear and a smooth evolutionary transition to empathy.

Alike the MAR, the PAM relies on salient cues to activate the adequate representations, based on a largely automatic process without any conscious input. Preston and de Waal also insisted

[8] Though we could imagine that in human cognition, cognitive effect could exist as a possible type of response, but it would alongside all others.

on the unlikeliness that target and observer's state exactly match in cases of empathy. The matching can be more or less accurate, generally in virtue of similar representations in both sides, but never an absolute match. The described nervous design creates an *appearance* of reciprocity by automatically activating an empathic response when it is needed for survival. As activated representations are the observer's, empathy is stated to *always* be a projection, of own's own representations to another individual's state (Sperber & Wilson 1995). Maybe this is the key to understanding Fabb's (2021) "representation matching mechanism" described in the previous section: the matching is only apparent because adequate affective responses are automatically generated after perceiving affective cues in the target and attribute these responses as matching our perception. Seen this way, cognition does not aim for state matching at all, as it only maps cues with automatic inferences and responses for practicality. Cue-salience increases the chances that an event is attended, thus an observer will pay more attention to it and corresponding representations will be more activated. In relevance-theoretic terms, this would correspond to a manifestness increase of an array of implicatures in the observer's cognitive environment[9]. The more salient an event is, the more likely it will be attended. In turn, increased attention paves the way for activation of the perception-action circuits and provide opportunity for empathy. Based on developmental studies and autism-related "empathy disorders", infants have been shown to attend to adults' non-verbal cues very early and the imitation of those expressions plays a crucial role in learning how to regulate, recognise, and organise emotions. Indeed, imitation helps assigning those cues to a specific set of emotional responses, that is creating a representation. Akin to the Somatic Markers hypothesis of emotions (Damasio 1994), the PAM also postulates for representations being linked to feeling states. Additionally, cue-salience could also explain how manifestness work in relevance-theory, the more a cue is salient, the more their associated representations will be activated. The more a cue resembles to a cue of activation of representations, the more those will be activated at the expense of others less resemblant ones.

The PAM is compatible with the appraisal theory if we consider patterns of activation as the equivalent to patterns of appraisal: an appraisal is thus the combined responses of the representations automatically activated when perceiving an event. Following that reasoning, as we equated before that the processes underlying appraisals and inferences are the same, if the patterns of activation in the PAM are also of appraisals, those patterns of activation appear to be the ones Sperber and Wilson (2015) discussed. It makes possible to posit that appraisals

[9] The authors used the term "more activated" rather than a manifestness-related notion since the manifestness comes from relevance-theory. However, as their functional resemblance is striking, I use them as equivalent in this context.

are expressed the same way in relevance theory, expressing emotion elicitation as using patterns of activation involving automatic inferences. Overall, perception gives rise to activation of patterns, which automatically generate an associated response. In a state-matching paradigm, perception of an emotional state will activate the corresponding patterns in the observer, given her patterns are similar to the target's.

Akin to the MAR, The PAM accounts for empathy as a matter of pattern sharing, and also as a goal-oriented process. Interdependence means the target and the observer have common goals, interdependence entails a cooperation that can be "temporary and superficial […] for a local goal", or "long lasting and deep […] for long-term goals" (Sperber & Wilson 1995: 5). Empathic emotions motivated by personal goals is also reflected in humans by a larger propensity to empathise and help targets who can help the observer. This supports our proposal about the required goal connivance between the observer's and the target's apparent goals for an empathic response to arise. Interrelationship between target and observer will increase the observer's attendance to the target's event and increase the similarity of activated representations between both. This supports situational empathy and effects of familiarity/similarity, which increase the chances of sharing similar interpretations, and thus, increase the chances of a somewhat "matching" affective response. Interdependence and interrelationship might together parallel in some ways engagement. This is an interesting possibility to consider, as engagement is a factor for eliciting emotional transportation, a state in which emotions are much more prone to happen and much more intense just like with interdependence and interrelationship. It seems like shared goals and similar representations are intimately correlated in real-life, as kins or individuals of the same social group are more likely to share goals and representations with other members.

The PAM offers the most satisfying explanation for how metacognitive acquaintance works from Sperber and Wilson's view. As the change in the receiver's cognitive environment happens spontaneously from perceiving a change in the speaker's behaviour, this automatic cognitive modification is better explained by the PAM as an automatic activation of the corresponding representations. The activation in turn triggers responses, resulting in a change in the cognitive environment. In cases of empathy, those responses are mainly affective. The shift is even more natural that we reinstated metacognitive acquaintance as being representations, and relying on perceptual representations, which the PAM is about.

The PAM also sheds light on the way on the procedural meaning works in relevance theory. Indeed, we already evoked the relationship between cues in emotional non-verbal

communication (such as visual facial cues and gestures) and emotional procedures, and additionally, Jones (2015) claimed that classifiers in sign language do not encode conceptual meaning, but procedural meaning orienting the receiver's interpretation construction. It points to perceptual cues being intertwined with procedures. We can also see another parallel between procedural meaning and Wilson and Sperber's (2015) indeterminate cases of communication which both are associated with making manifest a wide array of weak implicatures (2015: 123)[10]. Let us take their example (Sperber & Wilson 1995: 55) of a case of non-verbal and indeterminate communication: Mary and Peter are by the sea, Mary sniffs ostensively in presence of Peter, communicating something so loose and indeterminate and Peter cannot really pin down her intention; we conventionally call this import an impression. The way non-verbal cues in this example (e.g., sniffing) overlaps with how procedural meaning functions with other perceived cues, and how they are both usually associated with making manifest a wide array of weak array implicatures, could indicate that those two phenomena describe a similar functioning.[11] The PAM would account for all these cases using the notion of cue salience. Indeed, cue salience "differentially activates representations" (Preston & de Waal 2002: 16) and "increases the likelihood and the extent to which a representation is activated" (2002: 18), by increasing the chances of attending the salient event. The relation here with the determinate-indeterminate continuum is that cues in both views trigger an array of contents to varying degree depending on how much those cues stand out or competing with one another in a context. If cue interpretation necessarily involves activation of procedures, then procedural meaning can be regarded as a wide array of representations, each one possessing a similar cue of activation, that are activated on attending a given cue. The degree of activation of a representation depends on the similarity between the presented cue and the cue of activation, comprised in the notion of salience. It could be argued that this proposal does not explain emotional procedures from emotions themselves, because emotions are not perceived in the way perceptual cues are. An answer could be that the term *cue* must be taken in a in larger sense which includes perceptual cues about the internal feeling of affects. This means that a response-output (e.g., an affective feeling) generated by activating a representation could constitute in turn a cue-input for another array of representations with that specific response as their cue of activation (i.e., all representations with that specific affective feeling as a cue of activation). This includes affective responses which activate an array of representations with this specific affective response as their cue. For example, representations activating an *anger response,* alongside changing the affective state and produce the feeling of anger, will in turn

[10] Whereas more determinate cases of communication are associated with a narrower array but of more strongly manifest implicatures.
[11] Another possibility would be to say that sniffing is understood through procedural meaning, however involving other types of procedures.

activate another array of representations possessing the *anger response* as their cue of activation. This is consistent with Piskorska (2018), who described emotions in the relevant-theoretic framework "as factors prioritizing access to chunk of information in cognitive processes" (2018: 109) and "act as factors facilitating access to some contextual assumptions rather than others, out of a set of contextual assumptions that the speaker intends to make weakly manifest, and thereby, affecting the interpretation" (2018: 110). It supports a view in which representations with the same emotional cue of activation are activated all at the same time and acting as emotional procedures. Furthermore, she also presented a similar example to the one above: "an occurrence of an emotion facilitates recollection of all related assumptions that cooccurred with that emotion in the past, thus, experiencing happiness will make us notice happy things more easily" (2018: 106). And she also briefly evoked that we could assume that "an affect is a response to a set of previously derived representations" (2018: 110) corroborating the PAM perspective of patterns of activation. We could even imagine certain representations having motor cues, or more *abstract* cues, such as imagined ones (which was proposed by Preston & de Wall 2002: 11) or mentally represented concepts. This comparison points procedural meaning as potentially being a more widespread phenomenon in our inferential processes and cognition, as a way to chain representations with one another, extending inferences, and form longer processing pathways.

The PAM and the MAR, share many aspects in common while being grounded in different theoretic backgrounds. They might indeed describe parts of the same phenomenon from slightly different angles, one from the representational aspects of actions, the other from representational side of inferences. The PAM nicely complements the MAR by explaining how activation of representations in relevance theory corresponds to affective responses. Conversely, the MAR can explain how cues activate procedures to retrieve the corresponding representations via relevance-yielding processes and procedural meaning, ultimately leading to effects. As stated before, if we equate patterns of activation to patterns of appraisals, then the resulting interpretation of the MAR process can be considered an appraisal. Furthermore, degree of activation ties up closely with manifestness in relevance theory.

The PAM model offered an elegant phylogenetic continuum on how empathy emerged from a more general mechanism. Then, this global mechanism is refined to include representations involving others' mental states. This view on the evolution of empathy and affective processes parallels nicely with Wharton and Cornell's (2021) account on how human cognition. They explained that mechanisms of relevance found in human cognition can be explained with older

mechanisms incremented with new mechanisms added thereafter. The old systems remain and are not replaced, and the new ones add allow for new operations and enhance the capabilities of old systems.

On the input side, cue-recognition cannot be rendered propositionally since a cue is more akin to percepts. We can account for this in relevance theory through impressions or procedural meaning, pointing towards and making manifest a wide array of implicatures. However, Fodor (2007) proposed the notion of unconceptualized representations as a way to talk about perceptual content that is *represented*, but not yet *represented as* (i.e., not yet recognised and conceptualised for formal inferences). To support their existence, he took the example of the *echoic buffer* from Sperling's (1960) work, a memory space in which perceptual information is stored raw in their sensorial format (similar to a spectrogram for sounds or a visual map for images). He intended to show that there are iconic representations from which perceptual inferences can take place. Perceptual inferences and representations in the PAM work exactly that way, and it shows there might be nonetheless aspects of relevance which might involve thoroughly non-propositional contents.

On the output side, more interestingly, as the PAM focuses on action and (affective) responses, it implies that non-propositional effects are an alternative outcome of inferential alongside cognitive effects. In the end, assuming that the MAR and PAM tackle the same representations, such as in cases of empathy (and affective responses in general), it broadens the possibilities of relevance-theory and the range of its cognitive effects, by including somatic and autonomic responses as the result of automatic inferential processing.

The overlap between the representations in the MAR and the PAM and the notion of *response* strongly imply the notion of cognitive effect must be reworked to include other kinds of non-propositional effects. Since the PAM representations links cues to responses, neither of the those is properly conceptual content as it stands but is nonetheless a kind of automatic inference. Sperber and Wilson (2015) explained those mechanisms forming impressions as "activation or inhibition caused by brain states that represent information in all kind of ways" (2015: 137), potentially involving "unconscious weightings of feature" (2015: 137) and "fulfil the function of making some information available for processing" (2015: 137) but are still "genuinely inferential" (2015: 137). The PAM can explain at the level of representations the emergence of impressions as an automatic activation of representations, and the MAR shows that this account is compatible with a relevance-theoretic account. There is clearly room for all to integrate many underexplored aspects of cognition into relevance theory.

6.4. Perceptual symbol systems and non-propositionality

Barsalou (1999) proposes the perceptual symbol systems to describe inferential processes in relation to perceptual states. He proposes a way to explain propositionality with non-propositional components. This is essential to the MAR, as it offers to reconcile the non-propositional representations and propositions as is used within the relevance-theoretic framework. In this view, cognition and perception share a similar functioning, positing that cognition is inherently perceptual. Their functioning is opposed to conceptual systems. As Barsalou (1999) explains, those systems historically separated perceptual contents from cognition in favour of an amodal symbol system, manipulating symbols through logical derivations. It is called amodal because the internal structure of symbols is arbitrarily assigned to its referents. Thus, the internal structure holds no resemblance with its referent. In other words, the acquisition of conceptual representations implies an arbitrary link to the perceptual states that activate them, and both remain distinct in kind. Amodal systems are essentially "amodal because their internal structures bear no correspondence to the perceptual states that produced them" (Barsalou 1999: 578). As such, they imply that cognition and modalities of perception are distinct from each other and operated by separate systems. Thus, representations in amodal systems are non-perceptual and their relation to the referent is arbitrary. However, Barsalou points out there is little direct empirical evidence supporting the existence of amodal symbols in the mind. Furthermore, amodal systems are very potent at explaining phenomenon *post-hoc* but very limited in making predictions *a priori*. They cannot explain how symbols are grounded and do not integrate well in other fields such as neuroscience and perception. Finally, they are lacking in capability to explain how we represent spatio-temporal knowledge, relying on convoluted and unwieldy axioms (Barsalou 1999) (see the simulation theory sub-section, with Oatley and Johnson-Laird's (2021) example on spatial inferences).

The MAR with its two types of representations relying on perceptual and experiential cues also rejects a fully amodal view of cognition. The amodal view separating cognition from perception is also the one traditionally employed within relevance theory. It follows that relevance theory shares the issues evoked in the last paragraph as a result of using propositions for inferences. Impressions and procedurality are notions initially conceived to bypass perceptual modalities. Both change the manifestness of an array of implicatures, which remain propositional content. Even in mental imagery, ultimately, "images may help the increase manifestness of an array propositions" and produce implicatures (Wilson & Carston 2019: 37). In the end, amodal propositions are the backbone of relevance theory and that is why we needed new notions to account for affects and emotions to develop the MAR.

Perceptual symbol systems suggest a different conception of inference and can solve all the mentioned issues. During perception, sensorimotor information from the environment and the

body is parsed into perceptual properties. Only some selected aspects of perception are represented. The represented information is stored in systems which can be reactivated later. The brain stores this information in the form of a perceptual symbol, which is "a record of a neural activation that arises during perception" (Barsalou 1999: 582). As a result, Barsalou's theory might be useful to establish continuity between non-propositional representations and propositions.

Barsalou (1999) states that "unconscious mental representations – not conscious mental images – constitute the core of perceptual symbols" (ibid.: 583). So, symbols are representations in nature. Consequently, what is said for symbols also applies to the representations in the MAR. More accurately, perceptual symbols are representations containing a schematic aspect of a perceptual state. In this view, stored representations are organised in connectionist networks and are activated dynamically in response to situations and activated in different manners from the original situation of their creation. Different contexts change the patterns of activation, as contextual features bias activation towards some elements of the pattern more than others. Symbols are organised into simulators, allowing one to build limitless simulation of events and entities even if they are not present. Simulators allows for productivity, a requirement to form propositions.

Simulation of events is possible, because perceptual symbols are systematically organised to produce a coherent experience of the event on extraction. Simulations will never be complete, as only certain aspects of the event are selected. It follows that simulations tend to be biased and non-veridical as they come from perception and a specific point of view. Furthermore, heuristics and other processes distort further our representations of events by adding further information to what is being perceived. Mechanisms with genetic constraints such as emotion, space or movement processing play a central role in establishing, maintaining, running and organising simulators. As a result, simulators are regarded as both *empirical* and *rational* systems (Barsalou 1999). In virtue of emulating events and processes, simulators have a strong parallel with the simulation theory and mental models. Barsalou stated that mental models are indeed very similar to simulators.

According to Barsalou, perceptual symbols "represent schematic components of perception, not entire holistic experiences" (1999: 582). They are multimodal, originating from "sensory modalities, proprioception and introspection" (ibid). Related perceptual symbols are organised into "a simulator that produces limitless simulations of a perceptual component" (ibid). What organise symbols within a simulator are frames. A frame is "an integrated system of perceptual symbols that is used to construct specific simulations of a category" (ibid: 590). A frame plus the simulation it produces form a simulator. Barsalou (1999) exemplifies frames using the

different conceptualisations of *foot*. Depending on the context, we simulate *foot* differently, such as the foot of a tree, of a human or of a horse. It follows that different symbols will be accessed in different contexts surrounding *foot* or any perceived event, and they are accessed thanks to the frame corresponding to each conceptualisation. These interactions between frames and conceptualisations means that words associated with simulators allow the construction of simulations to be controlled by linguistic means.

Perceptual symbols are multimodal. They can represent aspects of experience originating from sensorial sources, such as audition with speech and sounds, or touch with temperature and textures. They can originate from proprioception with movements and body positions. More interestingly for this work, they can also originate from introspection with representational states, cognitive operations, and emotional states. It means that selective attention can focus on a specific introspective aspect of our emotions, moods and affects and store it for later as a symbolic representation. Those three sources of multimodality offer an interesting parallel to the perceptual and metacognitive representations involved in metacognitive acquaintance in the MAR. Perceptual representations are related to senses and proprioception. They associate aspects of (external or internal) perceptual experience with a response. Metacognitive representations are related to introspection. They associate a mental state with a response. As a result, it could be argued that metacognitive representations focus on aspects of our mental experiences (i.e., changes in our cognitive environment). Specific aspects of those changes may be selected by relevance yielding processes, stored in the form of a representation, and later be reactivated to trigger inferences when similar changes reoccur. Thus, Barsalou (1999) links cognition to perception with representations in a similar way to the PAM (Preston & de Waal 2002) and the MAR. However, perceptual symbols are representations of schematic perceptual states whereas representations in the PAM links a cue to a response. A cue is the closest notion we have to symbols, but there is no clear equivalent to responses in perceptual symbol systems. A possibility may be to envision responses as the resulting inferences realised by perceptual symbols. Those two views are compatible if we assume that representations in the PAM encompass the same thing as perceptual symbols. It implies that responses in the PAM are constituents of simulations and bear some of the characteristics of sensory-motor experiences from which they are derived.

Barsalou (1999) explains linguistic symbols as inherently perceptual representations that are not different from perceptual symbols. That is, a perceptual symbol represents the schematic memory of a perceived event in general. A linguistic symbol represents the schematic memory of a perceived event that is a spoken or written word. As linguistic symbols become associated

to simulators, so, words can control simulations. Thus, language "provides a powerful means of constructing simulations that go far beyond an individual's experience" (ibid: 592). It becomes easier to explain why narratives can be considered as simulations ((Mar et al. 2011; Mar & Oatley 2008; Oatley 2016) or allow people to experience situations they would normally not be able to (Oatley 2016, Mar et al. 2011).

Barsalou (1999) presents four properties of symbols allowing them to fully account for concepts on a non-propositional basis: productivity, propositionality, variable embodiment, and conceptual abstractness. The first property is productivity. Simulators can be "combined combinatorially and recursively to implement productivity" (ibid: 582). Therefore, an unlimited number of simulations can be generated from a limited number of representations using combination and recursion mechanisms.

The second property is propositionability. Simulators can be bound to perceived individuals necessary to integrate propositions, just like concepts can refer to objects in the real world. In fact, Barsalou equates simulators with concepts. "[P]ropositions involve bringing knowledge to bear on perception, establishing type-token relations between concepts in knowledge and individuals in the perceived world." (Barsalou 1999: 595). As such, perceptual symbol systems possess all the necessary properties of amodal systems. They can represent true or false propositions depending on whether a simulation succeeds or fails/ is absent. They produce alternative interpretations of the same event. Perceptual simulations can also be paraphrased using alternative simulators. They involve token-type mappings of symbols with construed individuals. And finally, perceptual simulations can be productively assembled to form complex hierarchical propositions. All in all, perceptual symbol systems *are* propositional systems.

Subsequently, Barsalou suggests that this allows to see a continuum between animal and human cognition as well as between infants and adult cognition. Unifying human cognition with other types is better achieved if we envision a system primarily relying on non-propositional representations producing propositional structures. It is made possible thanks to additional mechanisms allowing one to reflect metacognitively on our own mental states to process simulators recursively in a hierarchical way. We are not saying that human cognition is identical to animal cognition. We are just underlining that both kinds of cognitive processes may rely on a similar chassis. Human cognition can be described as new systems added on top of older ones to allow for recursion, language, and metacognitive representations. Consequently, for the MAR, non-propositional representations may be the building blocks of propositions. They are not antithetical to or distinct from one another. It paves the way for a

possible continuum from not-propositional-like to propositional-like structures. The more a pattern of activation is propositional in its properties and in the way it is structured, the closer it is to the propositional-like end of the spectrum.

The third property is variable embodiment. As perceptual symbols are tied to sensory-motor systems, they are embodied in the sense that a "symbol's meaning reflects the physical system in which it is represented" (Barsalou 1999: 598). Barsalou addresses concept stability by proposing that individuals share many factors in the acquisition of simulators, such as "a common cognitive system, common experience with the physical world, and socio-cultural institutions that induce conventions" (1999: 588). His point is that, even though a concept BIRD can be conceptualised differently between individuals and within an individual across contexts, it acquires stability if individuals can simulate others' conceptualisations. Barsalou quoted an unpublished study showing that each subject could produce a unique conceptualisation, and still accepted others' conceptualisations vastly different from theirs. As a result, concept stability stems from being able to simulate others' conceptualisations. Furthermore, in communication, participants are constrained by a common context, driving their simulations to be even more similar. Despite individual differences, it is possible to have *shared embodiment*, that is, humans share approximately the same apparatus. It means that mechanisms used to perceive the colour red will represent it approximately in the same say across all humans. Thus, we will have approximately the same conceptual representations of basic perceptual properties.

The fourth property is conceptual abstractness. Complex simulations allow perceptual symbols to represent abstract concepts if those simulations involve a combination of physical and introspective elements. Abstract concepts can thus be represented following three steps (1999: 603):

(a) identifying the event sequences that frame them,

(b) specifying the physical and introspective events in these sequences, and

(c) identifying the focal elements of these simulations.

Barsalou demonstrated that abstract concepts are constituted by assembling more concrete simulations together. What makes abstract concept seem non-perceptual is that they strongly rely on "complex configurations of multimodal information distributed over time" (1999: 603). In other words, even abstract concepts are perceptually grounded.

The perceptual symbol systems help further the integration of non-propositional content into our current understanding. While it is not clear how the representations in the MAR with the

notions of cue and response exactly map onto perceptual symbols, it still offers an argument in favour of representations accounting for propositions. Indeed, representations in the MAR can be seen as working similarly to perceptual symbol systems, and thus can also produce concepts, simulations, and propositions. Both systems involve representations about multimodal aspects of perceptual experience, including inner mental states. Both can be conceptualised in terms of networks of simulative experiences, since in the MAR, a response from a representation can activate another representation associated with that response. It turns out properties of Barsalou's perceptual symbol systems could be transferred to the MAR to account for propositions.

In return, the relevance theory can address some of the shortcomings of perceptual symbols. Mainly, perceptual symbols do not specify how attention selects certain features but not others, how do we compute from one representation to another, or how combinations of concepts are constrained. All these three issues can be explained by the cognitive principle in relevance theory. For the first issue, we selectively attend to features that optimise cognitive effects to increase the degree of relevance. For the second issue, the activation of a representation changes the manifestness of other relevant representations it interacts with, creating implicatures. The MAR explains the chain of events between representations in terms of responses (or effects) activating other representations with a cue of activation corresponding to the response. For the third point, concepts are selected when they are relevant enough. Concepts could be modified, for example, following the process of ad-hoc concepts construction (Wilson & Carston 2007). In this process lexical adjustments occur following the rules of relevance, resulting in narrowing, broadening or metaphorical extension of the encoded meaning. Perceptual symbol systems offer an interesting way to tackle propositionality. Thus, propositionality should be viewed as an "emergent property" of the interaction between non-propositional representations and how their network structure them together hierarchically.

6.5. Emotion as the inferential output

The MAR shows that the outcome of interpretation is not only cognitive effects but can be an affective state as well. Indeed, Piskorska (2018) suggests that along with a presumption of positive cognitive effects, there should also be a presumption of experiencing an affective state in some situations. An affective state leads to positive affective effects (de Saussure & Wharton 2020), by making accessible certain contextual assumptions over others. More interestingly, Piskorska suggests that the emotions can be an integral part of the interpretation, alongside other retrieved positive cognitive effects thanks to the emotion. This view of *emotion*

as information is probably what happens in the empathic process: the result of empathy is experiencing an affective state (that we can attribute to be the same as the target's). In some cases, feeling the state can also be the ultimate goal of interpretative processes, and this without considering the cognitive effects potentially yielded on the basis of the current affective state in the next inferential 'loop'. Hatfield et al. (1994) hold similar positions regarding emotional contagion, that "instances of emotional communication convey not only conceptual information about emotional states, but also ultimately and above all, something of '*the emotional states themselves*' (1994: 201). De Saussure and Wharton (2021) evoke convergent views with their notion of affective effects activating past experiences and associated feelings (namely in literary contexts), within the picture of a potential *affective relevance*.

Feelings as the ultimate purpose of a cognitive activity are especially visible in appreciating art, music, and literature, where a substantial part of the enjoyment comes from attaining these affective states. The experience is at the core of these activities, and not always are cognitive effects the desired outcome of readers of a book. Cases of phatic communication may also fit into this frame, because what is aimed here is not necessarily additional cognitive effects but to establish a contact, or maybe experience positive emotions linked to socializing and maintaining relationships. This is consistent with the PAM, in which the purpose of cognition is not necessarily to draw true conclusions about the world (that is, implicatures and cognitive effects) but to elicit responses adapted to a situation. Those responses entail feelings, emotions, and sensations as autonomic reactions, all of which can in turn influence inferential processes and favour somatic behaviours. Though it may be possible that in human cognition implicatures (propositional) help deliver a deeper, more accurate, and more economical description of the world, it is naturalistically dubious to think that they are disconnected from responses and feelings, commonly regarded as prevalent in animals. Envisaging implicatures as having inherent behavioural components tied to them may bear some advantages, such as better accounting for the phylogenetic continuum between animals, apes, human cognition as well as infant and adult cognition. In other words, our inferential abilities make more sense if they are ultimately designed to produce non-propositional effects in the form of feelings, affects and actions with propositional effects as a means to this end.

7. Conclusion

In this work, we examined how empathy and narrative empathy are encompassed in different fields. Drawing upon the mechanisms, predictions and outcomes in the appraisal theory and cognitive narratology, we built a model using relevance-theoretic notions, the Model of Affective Relevance. Empathy is taken here as one possible outcome of a larger mechanism rather than a standalone mechanism. In the analysis, what differs empathy from narrative empathy is a matter of goals relevant to the current situation. Narrative empathy requires narrative goals, but affective goals can still interact with appraisal and engagement in the character. The range of applications of the MAR extends much further than narratives and empathy, as it argues that empathy arises for real people in a very similar way as for fictional characters. It provides explanation to cases of sympathy or in which no affective reaction is observed and offers insights into how we understand creative uses of language, artworks, and other people's minds.

I propose that Wilson and Sperber's notion of metacognitive acquaintance is a key to understand empathy. We extended this notion to incorporate representations involving a logical association between a cue and a mental state. Metacognitive acquaintance relies on two subsequent levels of representations to work with: firstly, perceptual representations to account for automatic changes in one's cognitive environment, and metacognitive representations to attribute mental states by identifying these changes. Accessed via procedural meaning, metacognitive representations allow for mentalistic inferences. In a narrative context, metacognitive acquaintance is crucial for appraising character and events, and it is necessary for eliciting character-driven emotions. Metacognitive acquaintance is essentially all the knowledge we have of one's own and others' mental variables, including goals, intentions, emotions, feelings, personality, desires, and affective responses, all constructed through our inferences and experiences of those states.

The MAR will be useful to account for emotion elicitation and affective responses within relevance theory. The discussion generated might provide useful insights on the accountability of non-propositionality and offer ways to integrate relevance theory within a phylogenetic perspective, to perhaps observe a continuum from the non-propositional to the propositional, and an extension of the scope of cognitive effects.

One of the main challenges for the MAR is to integrate other non-propositional effects alongside affective effects, such as perceptual, sensorimotor, and poetic effects. Hopefully, this can accommodate for other aspects of cognition by changing the type of effects arisen from perceptual representations, changing the attributions made on the basis of the change. An idea for future research is accounting for mental imagery (Wilson & Carston 2019) using our model with perceptual effects, in the same as our model account for empathy with affective

affects. As mental imagery involves a simulation of perception-like images, it could be that this simulation shares a functioning similar to the simulation of other minds during perspective-taking.

The large grey area of this model is the definition of goals from a representational and processing standpoint. We acknowledge that 'goal' is an umbrella term for many motivational factors. However, we have not addressed whether they are propositional or non-propositional, that is, their exact nature as is used in relevance theory. One direction for future research is to explore the properties of goals, their relationship with representations, and how they interact with relevance. My hypothesis is that any representation or an array of representations can be taken as a goal under the right circumstances, as long as they generate pleasant or unpleasant affective responses tied to implicatures during inferential processes. That is to say, that goals are representations that produce a feeling, with a positive or negative sensation counterpart constituting the motivational drive. Establishing goals as representations give rise to an affective response as a result of the inferential process. Goals may be seen as cognitive highways. Just like engagement and representations could be taken as both a process and a structure, goals share this duality if we consider goals as made of representations. The activation a large and interconnected array of representations leads leading to an array of pleasant and unpleasant feeling outputs. Combinations of those feeling outputs correspond to appraisals being essentially patterns of activation. When the joint outputs are sufficiently salient to our attention in our cognitive environment, and we recognise those patterns of activation, we would call them emotions or affects. Further discussion is needed on the relation between goals, representations, and appraisals.

We also compared our proposed model with some other theoretic models and with similar conclusions and found that they complement each other in our understanding of metacognitive acquaintance and the role of emotions and feelings that arise in inferences. With the same mechanisms of conditioning, we can easily explain why we associate a cue with mental state.

Empathy is only the tip of the iceberg of explaining affective responses. There is an underlying common mechanism in cognition that can explain affective responses and mentalistic inferences, and it has the potential to go far beyond. The functioning of representations as described here may be extended to account for other areas of cognition such as perceptual simulations. That would be possible if some representations are involved in the production of specific kinds of responses, such as sensorimotor responses. The MAR may also provide insights into poetic effects and synaesthesia in literary works and artworks. This model is only

the first step towards a unified view of cognition. It will need further improvements to unveil what is quintessential to thinking.

8. Acknowledgments:

Warmest thanks to Mengyang Qiu, Tim Wharton and Chara Vlachaki for all their support and insights.

References:

Aarts, H., Gollwitzer, P. M., & Hassin, R. R. (2004). Goal Contagion: Perceiving Is for Pursuing. *Journal of Personality and Social Psychology*, *87*(1), 23–37. https://doi.org/10.1037/0022-3514.87.1.23

Aristotle. (1970). *Poetics*. University of Michigan Press.

Bailey, O. (2022). Empathy and the value of humane understanding. *Philosophy and Phenomenological Research*, *104*(1)

, 50–65.

Baird, A. D., Scheffer, I. E., & Wilson, S. J. (2011). Mirror neuron system involvement in empathy: A critical look at the evidence. *Social Neuroscience*, *6*(4), 327–335.

Bal, P. M., & Veltkamp, M. (2013). How Does Fiction Reading Influence Empathy? An Experimental Investigation on the Role of Emotional Transportation. *PLOS ONE*, *8*(1), e55341. https://doi.org/10.1371/journal.pone.0055341

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2), 241–251.

Batson, C. D. (1991). *The altruism question: Toward a social-psychological answer*. N.J.: Erlbaum.

Batson, C. D., Batson, J. G., Slingsby, J. K., Harrell, K. L., Peekna, H. M., & Todd, R. M. (1991). Empathic joy and the empathy-altruism hypothesis. *Journal of Personality and Social Psychology*, *61*(3), 413–426. https://doi.org/10.1037/0022-3514.61.3.413

Blakemore, D. (1987). *Semantic constraints on relevance*. Macmillan.

Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers* (Vol. 99). Cambridge university press.

Blakemore, D. (2011). On the descriptive ineffability of expressive meaning. *Journal of Pragmatics*, *43*(14), 3537–3550.

Bortolussi, M., & Dixon, P. (2003). *Psychonarratology: Foundations for the empirical study of literary response*. Cambridge University Press.

Bruner, J. S. (1986). *Actual Minds, Possible Worlds*. Plenum Press.

Carroll, N. (2001). *Beyond aesthetics: Philosophical essays*. Cambridge University Press.

Carruthers, P., & Smith, P. K. (1996). *Theories of theories of mind*. Cambridge university press.

Cave, T., & Wilson, D. (2018). *Reading Beyond the Code: Literature and Relevance Theory*. Oxford University Press.

Clore, G. L., & Ortony, A. (2008). Appraisal theories: How cognition shapes affect into emotion. In *Handbook of emotions, 3rd ed* (pp. 628–642). The Guilford Press.

Coplan, A. (2004). Empathic engagement with narrative fictions. *The Journal of Aesthetics and Art Criticism*, *62*(2), 141–152.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, *163*, 163–228.

Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. *Handbook of Emotions*, *2*(2), 91–115.

Craik, K. J. W. (1952). *The nature of explanation* (Vol. 445). CUP Archive.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience* (Vol. 1990). Harper & Row New York.

Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The Iterative Reprocessing Model: A Multilevel Framework for Attitudes and Evaluation. *Social Cognition*, *25*(5), 736–760. https://doi.org/10.1521/soco.2007.25.5.736

Damasio, A. (1994). Descartes' error: Emotion, rationality and the human brain. *New York: Putnam*, *352*.

Damasio, A. R., Tranel, D., & Damasio, H. C. (1991). Behavior: Theory and preliminary testing. *Frontal Lobe Function and Dysfunction*, *217*.

Davies, M., & Stone, T. (1995a). *Folk psychology: The theory of mind debate*.

Davies, M., & Stone, T. (1995b). *Mental simulation: Evaluations and applications-reading in*

mind and language. John Wiley & Sons.

Decety, J. (2010). The neurodevelopment of empathy in humans. *Developmental Neuroscience*, *32*(4), 257–267.

Djikic, M., Oatley, K., & Moldoveanu, M. C. (2013). Reading other minds: Effects of literature on empathy. *Scientific Study of Literature*, *3*(1), 28–47.

Dunbar, R. I. (2003). The social brain: Mind, language, and society in evolutionary perspective. *Annual Review of Anthropology*, 163–181.

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In *Handbook of affective sciences* (pp. 572–595). Oxford University Press.

Fabb, N. (2021). Experiences of ineffable significance. *Beyond Meaning*, *324*, 135.

Fabb, N. (2022). *A Theory of Thrills, Sublime and Epiphany in Literature*. https://anthempress.com/

Feagin, S. L. (2018). Reading with Feeling: The Aesthetics of Appreciation. In *Reading with Feeling*. Cornell University Press. https://doi.org/10.7591/9781501721465

Fodor, J. A. (1975). *The language of thought* (Vol. 5). Harvard university press.

Gallese, V. (2003). The roots of empathy: The shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology*, *36*(4), 171–180.

Gallese, V., Gernsbacher, M. A., Heyes, C., Hickok, G., & Iacoboni, M. (2011). Mirror Neuron Forum. *Perspectives on Psychological Science*, *6*(4), 369–407. https://doi.org/10.1177/1745691611413392

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, *8*(9), 396–403. https://doi.org/10.1016/j.tics.2004.07.002

Gerrig, R. J. (2018). *Experiencing narrative worlds: On the psychological activities of reading*. Routledge.

Green, K. (1993). *Relevance theory and the literary text: Some problems and perspectives*. *22*(3), 207–217. https://doi.org/10.1515/jlse.1993.22.3.207

Green, M. C., & Brock, T. C. (2002). In the mind's eye: Transportation-imagery model of

narrative persuasion. In *Narrative impact: Social and cognitive foundations* (pp. 315–341). Lawrence Erlbaum Associates Publishers.

Grice, H. P. (1957). Meaning. Reprinted in. *Studies in the Way of Words*, 213–223.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). Emotional contagion: Cambridge studies in emotion and social interaction. *Cambridge, UK: Cambridge University Press. Errors-in-Variables Regression Model When the Variances of the Measurement Errors Vary between the Observations. Statistics in Medicine*, *21*, 1089–1101.

Hoffmann, R. (2000). Twenty years on: The evolution of cooperation revisited. *Journal of Artificial Societies and Social Simulation*, *3*(2), 1390–1396.

Hogan, P. C. (2003). Cognitive science. *Literature, and the Arts: A Guide for Humanists*.

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual Review of Psychology*, *60*(1), 653–670.

Johnson, D. R. (2012). Transportation into a story increases empathy, prosocial behavior, and perceptual bias toward fearful expressions. *Personality and Individual Differences*, *52*(2), 150–155.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.

Johnson-Laird, P. N., & Byrne, R. M. (1991). *Deduction.* Lawrence Erlbaum Associates, Inc.

Johnson-Laird, P. N., & Oatley, K. (2021). Emotions, simulation, and abstract art. *Art & Perception*, *9*(3), 260–292.

Jones, S. (2015). Classifier constructions as procedural referring expressions in American Sign Language. *Research in Language*, *13*(4), 368–391. https://doi.org/10.1515/rela-2015-0032

Keen, S. (2006). A Theory of Narrative Empathy. *Narrative*, *14*(3), 207–236.

Keen, S. (Ed.). (2010). *Toward a cognitive theory of narrative acts* (1. ed). Univ. of Texas Press.

Keen, S. (2014). Narrative Empathy. In *Handbook of Narratology* (pp. 521–530). De Gruyter. https://doi.org/10.1515/9783110316469.521

Keen, S. (2015). Intersectional narratology in the study of narrative empathy. *Narrative Theory Unbound: Queer and Feminist Interventions*, 123–146.

Keysers, C., & Gazzola, V. (2009). Expanding the mirror: Vicarious activity for actions, emotions, and sensations. *Current Opinion in Neurobiology*, *19*(6), 666–671.

Kolaiti, P. (2015). *The poetic mind: A producer-oriented approach to literature and art*. 22.

Kolaiti, P. (2019). *The Limits of Expression: Language, Literature, Mind*. Cambridge University Press.

Konijn, E. A., van der Molen, J. H. W., & van Nes, S. (2009). Emotions bias perceptions of realism in audiovisual media: Why we may take fiction for real. *Discourse Processes*, *46*(4), 309–340.

Koopman, E., & Hakemulder, F. (2015). *Effects of Literature on Empathy and Self- Reflection: A Theoretical-Empirical Framework*. 33.

Kuiken, D., & Douglas, S. (2017). Forms of absorption that facilitate the aesthetic and explanatory effects of literary reading. *Narrative Absorption*, *27*, 219–252.

Lazarus, R. S. (1991). *Emotion and Adaptation*. Oxford University Press.

Levin, H. S., Eisenberg, H. M., Eisenberg, C. D. of N. H. M., & Benton, A. L. (1991). *Frontal Lobe Function and Dysfunction*. Oxford University Press.

Lipps, T. (1903). Einfühlung, innere nachahmung und organenempfindungen. *Revue Philosophique de La France Et de l*, *56*.

Mar, R. A., & Oatley, K. (2008). The Function of Fiction is the Abstraction and Simulation of Social Experience. *Perspectives on Psychological Science*, *3*(3), 173–192. https://doi.org/10.1111/j.1745-6924.2008.00073.x

Mar, R. A., Oatley, K., Djikic, M., & Mullin, J. (2011). Emotion and narrative fiction: Interactive influences before, during, and after reading. *Cognition and Emotion*, *25*(5), 818–833. https://doi.org/10.1080/02699931.2010.515151

Miall, D. S., & Kuiken, D. (1999). What is literariness? Three components of literary reading. *Discourse Processes*, *28*(2), 121–138.

Nussbaum, M. C. (1998). Poetic justice: The literary imagination and public life. *Political*

Theory, *26*(4), 557–583.

Oatley, K. (1992). *Best laid schemes: The psychology of the emotions*. Cambridge University Press.

Oatley, K. (1999a). Why fiction may be twice as true as fact: Fiction as cognitive and emotional simulation. *Review of General Psychology*, *3*(2), 101–117.

Oatley, K. (1999b). Meetings of minds: Dialogue, sympathy, and identification, in reading fiction. *Poetics*, *26*(5), 439–454. https://doi.org/10.1016/S0304-422X(99)00011-X

Oatley, K. (2004). From the emotions of conversation to the passions of fiction. *Feelings and Emotions: The Amsterdam Symposium*, 98–115.

Oatley, K. (2016). Fiction: Simulation of social worlds. *Trends in Cognitive Sciences*, *20*(8), 618–628.

Oatley, K., & Djikic, M. (2018). Psychology of Narrative Art. *Review of General Psychology*, *22*(2), 161–168. https://doi.org/10.1037/gpr0000113

Pilkington, A. (2000). Poetic effects. *Poetic Effects*, 1–228.

Piskorska, A. (2018). *Cognition and emotions – jointly contributing to positive cognitive effects?* https://doi.org/10.31338/uw.9788323520450.pp.102-111

Preston, S. D. (2007). A perception-action model for empathy. *Empathy in Mental Illness*, *1*, 428–447.

Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, *25*(1), 1–20. https://doi.org/10.1017/S0140525X02000018

Pylyshyn, Z. (2003). Return of the mental image: Are there really pictures in the brain? *Trends in Cognitive Sciences*, *7*(3), 113–118.

Rickard, N. S. (2004). Intense emotional responses to music: A test of the physiological arousal hypothesis. *Psychology of Music*, *32*(4), 371–388.

Saussure, L. de. (2021). An experiential view on what makes literature relevant. *Beyond Meaning*, *324*, 99.

Saussure, L. de, & Wharton, T. (2020). Relevance, effects and affect. *International Review of*

Pragmatics, *12*(2), 183–205. https://doi.org/10.1163/18773109-01202001

Scherer, K. R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of New Music Research*, *33*(3), 239–251.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, *48*(4), 813–838. https://doi.org/10.1037/0022-3514.48.4.813

Soussignan, R. (2002). Duchenne smile, emotional experience, and autonomic reactivity: A test of the facial feedback hypothesis. *Emotion*, *2*(1), 52–74. https://doi.org/10.1037/1528-3542.2.1.52

Sperber, D. (1996). Explaining culture: A naturalistic approach. *Cambridge, MA: Cambridge*, *1101*.

Sperber, D. W., & Wilson, D. (1995). *D.(1986): Relevance: Communication and cognition*. Oxford: Blackwell.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. 142). Citeseer.

Sperber, D., & Wilson, D. (2015). Beyond Speaker's Meaning. *Croatian Journal of Philosophy*, *15*(2 (44)), 117–149.

Titchener, E. B. (1909). *Lectures on the experimental psychology of the thought-processes*. Macmillan.

Tooby, J., & Cosmides, L. (2008). *The evolutionary psychology of the emotions and their relationship to internal regulatory variables.*

Trotter, D. (1992). Analysing literary prose: The relevance of relevance theory. *Lingua*, *87*(1–2), 11–27.

Vignemont, S., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences*, *10*(10), 435–441. https://doi.org/10.1016/j.tics.2006.08.008

Wharton, T. (2016). That bloody so-and-so has retired: Expressives revisited. *Lingua*, *175–176*, 20–35. https://doi.org/10.1016/j.lingua.2015.08.004

Wharton, T., Bonard, C., Dukes, D., Sander, D., & Oswald, S. (2021). Relevance and emotion. *Journal of Pragmatics*, *181*, 259–269. https://doi.org/10.1016/j.pragma.2021.06.001

Wharton, T., & Strey, C. (2019). *Slave of the Passions: Making Emotions Relevant*. https://doi.org/10.1017/9781108290593.022

Wilson, D., & Carston, R. (2007). A unitary approach to Lexical Pragmatics: Relevance, Inference and Ad hoc concepts. In N. Burton-Roberts (Ed.), *Pragmatics* (pp. 230–259). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-349-73908-0_12

Wilson, D., & Carston, R. (2019). Pragmatics and the challenge of 'non-propositional' effects. *Journal of Pragmatics*, *145*, 31–38. https://doi.org/10.1016/j.pragma.2019.01.005

UNIVERSITÉ DE NEUCHÂTEL

Rectorat
Fbg de l'Hôpital 41
2000 Neuchâtel
Tél. +41 32 718 10 20
messagerie.rectorat@unine.ch

Pledge of Honour*

I hereby declare that I have read and understood the material explaining plagiarism and the prevention thereof provided by the University of Neuchâtel and that I understand the procedures of accurate citation and bibliographical practice.

I confirm that my paper is the result of my personal research and is solely my own work.

I affirm that any formulation, idea, research, reasoning or analysis borrowed from a third party is correctly and accurately indicated as such, clearly and transparently, and in such a way that the original source is immediately recognisable, in respect of citation techniques and the author's rights.

I am aware that not documenting source material or not citing clearly, correctly and completely constitutes plagiarism.

I am aware that plagiarism is considered a serious offence within the University and that any case of plagiarism can entail administrative sanctions and disciplinary consequences (including expulsion).

I have understood that in the case of plagiarism, the file will automatically be transferred to the Rector's office.

In light of the above, **I declare that I have not plagiarised, nor committed any other kind of fraud.**

Last name: Pozner First name: Ismaël

Course of study: Master in cognitive science Faculty: Faculty of Humanities

Place and date: Moutier, 29 July 2022 Signature: Ismaël P.

****

This form is to be filled in by each student writing a significant paper (especially a Bachelor's or Master's thesis) or a Doctoral thesis. It must be included with each paper submitted.

*The text of this form has largely been inspired by the Rector's office's directive 0.3 bis *Directive de la direction 0.3 bis*, entitled *Formulaire Code de déontologie en matière d'emprunts, de citations et d'exploitation de sources diverses*, of the University of Lausanne, April 23rd 2007, and adapted for the requirements of the University of Neuchâtel.