

Politeness, Relevance and Scalar Inferences

Diana Mazarella

This is the peer-reviewed version of the following article: Mazarella, D. (2015) Politeness, relevance and scalar inferences, *Journal of Pragmatics*, 79, 93-106, which has been published in final form at <https://doi.org/10.1016/j.pragma.2015.01.016>.

Recent behavioural studies in experimental pragmatics investigate the effect of contextual manipulations on the derivation of scalar inferences (e.g. *Not all X-ed* inferred from an utterance of ‘Some X-ed’). Among these, Bonnefon et al. (2009) and Feeney and Bonnefon (2012) suggest that scalar inferences are less likely to be derived in face-threatening contexts. Indeed, they even suggest that a face-threatening utterance of the form ‘Some X-ed’ can be interpreted as communicating that *All X-ed*. This paper argues that the experimental evidence provided so far is compatible with two alternative explanations of the empirical data: (i) face-threatening contexts block the derivation of scalar inferences, or (ii) in face-threatening contexts the scalar inference is in fact derived as part of the intended interpretation but is less likely to be accepted (as true). Drawing on the theoretical distinction between ‘comprehension’ and ‘acceptance’ of the communicated content (Sperber et al. 2010), the paper proposes an analysis of the results in light of Relevance Theory. In line with (ii), Relevance Theory predicts that in face-threatening contexts the scalar inference *Not all X-ed* may be derived as part of the interpretation of the utterance but consideration of the communicator’s ‘preferences’ (e.g. her concern to be polite/kind) may lead the hearer to judge the scalar inference to be probably false and so to reject it. In such a case, the hearer may go on to infer that the reality is that *All X-ed* but not attribute this to the speaker as part of the intended meaning of the utterance.

Keywords: scalar inference, face threat, politeness, Relevance Theory

1. Introduction

The phenomenon of ‘scalar inference’ has been the subject of considerable debate within pragmatic theory and, more recently, has led to extensive empirical work in experimental pragmatics. Scalar inferences are pragmatic inferences like (1b), which seem to be drawn on a regular basis across different contexts.

- 1 a. Some of the guests have arrived.
- b. Not all of the guests have arrived.

While the encoded (semantic) meaning of the scalar expression ‘some’ is compatible with ‘all’, (1a) is often taken to suggest that *not all* the guests have arrived. In general, scalar inferences seem to arise when the speaker uses an expression which is not the strongest one on an informativeness scale, like the following: < some, many, most, all > (Horn, 1972, 1984).¹

Much of the debate has focused on the role of context in the derivation of scalar inferences. The question is whether they arise independently of context or as a function of contextual features which happen to be shared by many contexts and thus give rise to the impression that the scalar inference is regular. On the one hand, default accounts (e.g., Levinson, 2000; Chierchia, 2004) conceive of scalar inferences as default inferences, which are lexically associated but defeasible if contextually inappropriate. On the other hand, ‘context-sensitive’ accounts (e.g., Grice 1975/1989; Sperber & Wilson, 1986/1995; Carston, 1990, 1998) suggest that scalar inferences are derived only when contextually appropriate. These competing accounts assign different roles to context: according to the former, context cancels inappropriate scalar inferences

¹ In what follows I use the expression ‘scalar inferences’ as opposed to ‘scalar implicatures’. This terminological choice, borrowed from Bonnefon et al. (2009), is neutral as to whether the scalar inference is to be thought of as arising at the level of the explicit content of the utterance or as an implicature.

which arise independently of contextual support, whereas, according to the latter, it triggers appropriate scalar inferences which otherwise don't arise.

Default and context-sensitive accounts give rise to different psycholinguistic processing predictions with regard to 'lower bound' contexts. Lower bound contexts are those in which "all that is relevant, or can be known, is the lower bound" of scalar expressions (Horn, 1984, p. 13). For instance, consider the following example:

- 2 A: Is there any evidence against them?
 B: Some of their documents are forgeries.

(Levinson, 2000, p. 51)

In the context set up by A's question, all that it is relevant to know for the purpose of the discourse (i.e. answering to A's question) is that *at least some* of their documents are forgeries (that constitutes sufficient evidence against them). When contextual manipulations make the scalar inference irrelevant, the expected speed of interpretation of scalar terms (e.g. 'some') varies according to the two models. Default accounts predict that the time taken to derive the interpretation would increase in virtue of the extra cognitive effort involved in cancelling the inappropriate scalar inference (which has arisen automatically). Context-sensitive accounts, on the other hand, predict the opposite: since the scalar inference (e.g. *Not all of their documents are forgeries*) is not relevant, it won't be derived; no additional processing cost is thus involved. Consistent with the context-sensitive accounts, Breheny, Katsos and Williams (2006) found that lower-bound contexts decreased the processing time of scalar terms (see also Katsos, 2008; Katsos, Breheny & Williams, 2005).

In a very interesting further development, Bonnefon et al. (2009) and Feeney and Bonnefon (2012) have suggested that other contextual manipulations, beyond those explored by Breheny et al. (2006), may prove to be equally fruitful for testing default and context-sensitive accounts of scalar inference. In particular, they propose that, along with lower-bound contexts, *face-threatening* contexts may make the scalar inference contextually inappropriate. Face-threatening contexts are contexts in which the face of the addressee, that is, his public image and positive identity, may be damaged. For instance, in a face-threatening context like (3), the addressee might take the use of 'some' to be a polite device adopted by the speaker in order not to hurt his feelings. If this is the case, Bonnefon et al. (2009) suggest, A won't interpret B's utterance as conveying that some *but not all* the guests at the dinner thought that he drank too much.

- 3 A: What impression did I make during dinner?
 B: Some thought that you drank too much.

(Bonnefon et al., 2009, p. 250)

Indeed, they even suggest that A might interpret B's utterance as communicating that *all* the guests thought that he drank too much.

The primary aim of this paper is to show the methodological importance of the theoretical distinction between comprehension and acceptance, which has been arguably neglected in the experimental literature on scalar inferences. Because comprehension does not require acceptance of the communicated content, it is important to distinguish between, on the one hand, what the addressee takes to be the communicated content of the utterance (that is, what the speaker meant) and, on the

other hand, what the addressee accepts as true. These may differ from (and even be incompatible with) each other in significant ways.²

In what follows, I discuss the implications of taking the distinction between comprehension and acceptance into consideration for the current literature on politeness and scalar inferences. I suggest that there are two alternative lines of explanation that can account for the available empirical evidence: (i) face-threatening contexts block the derivation of scalar inferences (as suggested by Bonnefon and colleagues), or (ii) in face-threatening contexts the scalar inference is in fact derived but is less likely to be accepted (as true). I discuss the relevance-theoretic account of scalar inferences and suggest that this approach favours the explanation in (ii). Finally, I show that my proposal is also compatible with the effect of self-rated honesty on the derivation of scalar inferences, which is another component of the investigation by Feeney and Bonnefon (2012).

2. Politeness and scalar inference: experimental studies

Bonnefon et al. (2009) and Feeney and Bonnefon (2012) investigate whether face-threatening contexts behave like lower-bound contexts in blocking the derivation of scalar inferences. This amounts to testing whether politeness should be inserted among the list of contextual factors whose manipulation might be fruitfully applied in assessing alternative theories of scalar inferences (along the lines of Breheny et al. (2006)).

A face-threatening context is one in which the face of the addressee may be damaged, where an individual's 'face' corresponds to his/her public self-esteem, a property/quality which is cultivated (and defended) in social interactions (Brown & Levinson, 1987). The hypothesis put forth by Bonnefon et al. (2009) is that an utterance of the form 'Some X-ed' (e.g. "Some thought you drank too much") is less likely to be interpreted as communicating the scalar inference *Not all X-ed* when X-ing is something that threatens the face of the listener (e.g. thinking that the listener drank too much). In these circumstances, Bonnefon et al. suggest, the scalar expression 'some' may be interpreted as a polite device to sugar-coat a face-threatening act. The speaker's face-saving concerns might motivate her to use the term 'some' even if she is in an epistemic position to use the term 'all'. As a consequence, the hearer might refrain from concluding that the speaker does not know the stronger (and more face-threatening) statement to be true. For example, in (3), 'some' might be interpreted as a polite device employed by the speaker B in order to limit the degree of hurt or damage to the feelings of the addressee A, as compared to that of a stronger term pertaining to the same issue (as in, e.g., the more face-threatening utterance "All the guests thought you drank too much").

- 3 A: What impression did I make during dinner?
 B: Some thought you drank too much.

If this is the case, Bonnefon et al. claim, the scalar inference *Not all of the guests thought you drank too much* is inappropriate and its derivation blocked.

² An anonymous reviewer has drawn attention to a widespread equivocation in the literature between two distinct uses of the expression 'deriving an inference': (i) deriving an inference as a component of the interpretation intended by the speaker, and (ii) deriving an inference about the way the world is. This is precisely what is meant to be captured by my distinction between comprehension and acceptance as two stages in the process of forming beliefs via testimony.

In order to test this hypothesis, Bonnefon et al. (2009) ran a series of three off-line studies, in which behavioural responses were compared across face-threatening and face-boosting contexts. In the first experiment, participants were presented with two stories, one in a face-boost condition, and the other in a face-threat condition. The following represents a sample story:

- 4 Imagine that you have joined a poetry club, which consists of five members in addition to you. Each week one member writes a poem, and the five other members discuss the poem in the absence of its author. This week, it is your turn to write a poem and to let the others discuss it. After the discussion, one fellow member confides to you that “Some people **hated/loved** your poem”.

(Bonnefon et al., 2009, p. 251)

After reading each story, participants were asked to answer the following Yes/No question: ‘From what this fellow member told you, do you think it is possible that everyone hated/loved your poem?’ The percentage of negative answers was compared across the two conditions.

The results show that 42% of participants answered ‘Yes’ when asked whether it was possible that everyone hated their poem when told that ‘some’ did (as opposed to 17% in the face-boost condition). In other words, participants were more likely to think that everyone hated their poem when told that some did, than to think that everyone loved their poem when told that some did.

In the second experiment, a variant of the first one, participants were required to rate, on a 10-point scale anchored at *totally unlikely* and *totally likely*, how likely it was that the speaker would have used the word ‘some’ if she knew that the number of people who hated x /loved x was in fact 1. The question was repeated for all numbers up to 6 (where 6 corresponds to *all*). The membership function of the number 6 (i.e. to what extent does *all* enter in the concept of ‘some’?) was significantly higher in the face-threat condition (5.6/10) than in the face-boost condition (4.1/10). This suggested that the quantity 6 (i.e. all) in the face-threat condition was considered as more representative of the concept denoted by ‘some’ than the quantity 6 (i.e. all) in the face-boost condition.³

Overall, the results of these two experiments were taken to show that when X threatens the face of the listener, “then ‘some X -ed’ is less likely to be interpreted as implying ‘not all X -ed’” (Bonnefon et al., 2009, p. 254). Bonnefon et al. (2009) suggest that in face-threatening contexts the scalar inference from ‘some...’ to ‘not all...’ is less likely to be derived.

Finally, in the third experiment, participants were presented with four stories (two in the face-boost condition, two in the face-threat condition), where the epistemic state of the speaker was explicitly manipulated. In one version, the speaker was said to know that the proportion of people who X -ed was 3/6, in the second, 6/6. The following is a sample story for the 6/6 version:

³ In virtue of the addition of a Target Group condition, Experiment 2 shows that the effect of a face-threatening context is detectable only when the target of the face-threat is the listener (as opposed to a third party, one that neither the speaker nor the listener care about.) The two groups were inserted into the design of the experiment to control for the possibility that face-threatening contexts might simply correspond to lower-bound contexts (favouring ‘at least some’ interpretations), and face-boosting contexts to upper-bound ones (favouring ‘some but not all’ interpretations). The interaction effect ruled this possibility out.

- 5 Imagine you gave a speech at a small political rally. You are discussing your speech with Denise, who was in the audience. There were six other people in the audience. You are considering whether to give the same speech to another audience. Hearing this, Denise tells you “Some people hated/loved your speech”. Denise knows that 6 people out of 6 loved/hated your speech.

(Bonnefon et al., 2009, p. 254)

After reading each story, participants were asked to rate, on four different 10-point scales, how *accurate*, *considerate*, *honest* and *nice* it was of the speaker to use the word *some* in that context.

The ratings revealed that the use of *some* (e.g. “Some people hated your speech”) by a speaker who knows that everyone *X*-ed is considered inaccurate and dishonest in both face-threatening and face-boosting contexts (although the effect was slightly stronger in the former). However, in these circumstances, the use of ‘some’ was considered nice and considerate only if the context was face-threatening.

Overall, the results of the three experiments were interpreted as follows:

When *X* in ‘some *X*-ed’ threatens the face of the listener, individuals are less likely to infer that the speaker *meant* or knew that not all *X*-ed – and this is because they consider the possibility that the speaker might want to be nice more than to be precise. (Bonnefon et al., 2009, p. 255, *my emphasis*)

Interestingly, a similar pattern of results was found by Feeney and Bonnefon (2012) with regard to the scalar connective ‘or’. Adopting an experimental design similar to the one in Bonnefon et al. (2009), Feeney and Bonnefon (2012) concluded that the scalar connective ‘or’ is less likely to give rise to an exclusive interpretation (i.e. ‘A or B but not both’) in a face-threatening context (“You’ll take a pay cut or your vacation period will be reduced”) than in a face-boosting context (“You’ll get a pay-rise or your vacation period will be increased”). They interpret these results as strengthening the conclusion of Bonnefon et al. (2009). Given their similar experimental design, the methodological considerations presented in the next section are intended to apply to both studies.

3. Comprehension versus acceptance: A cognitive distinction and its methodological implications

3.1 Assessing the believability of speaker’s meaning

In this section, I suggest that there is an important theoretical issue which has significant methodological implications for the experimental paradigm adopted by Bonnefon et al. (2009) and Feeney and Bonnefon (2012) and, in fact, undermines their interpretation of their results. Specifically, I claim that the design of their studies misses the important theoretical distinction between what the addressee takes to be the communicated content of the utterance (i.e. the speaker’s intended meaning), on the one hand, and what he infers about the reality (the state of affairs in the world), on the other hand.

To begin with, let us introduce the distinction between ‘comprehension’ and ‘acceptance’.⁴ Such a distinction is implicit in Grice’s (1957) definition of speaker meaning as well as in the relevance-theoretic characterisation of ostensive-inferential communication (Sperber & Wilson, 1986/1995). In both cases, a communicator is said to display different layers of intention: the intention to inform the addressee about something (i.e. the ‘informative intention’, in relevance-theoretic terms) and the intention to make this basic intention overtly evident (i.e. the ‘communicative intention’, in relevance-theoretic terms). These intentions reflect the two main goals of communicators: to be understood (communicative intention), and to make their audience think or act according to what is to be understood (informative intention). Crucially, the communicative intention can be fulfilled without the corresponding informative intention being fulfilled. For instance, the addressee might comprehend the utterance without accepting the truth of what he has understood, that is, without ending up believing the communicated content of the utterance. As Sperber et al. (2010) suggest, the gap between comprehension and acceptance is typically bridged by ‘epistemic trust’: when the addressee assumes the epistemic and moral trustworthiness of the communicator, the recognition of her informative intention leads to its fulfilment.⁵

To illustrate the relationship between comprehension and acceptance, as well as the distinction between inferences that are part of the communicated content of the utterance and inferences that are not, let us consider the following example from Noveck and Sperber (2007). Jane and Henry are hosting a dinner party. The doorbell rings and Jane shouts at him from the living room:

- 1 a. Some of the guests have arrived.

In what follows I elaborate on three possible scenarios. In the first scenario the communicated content of the utterance is accepted as true, while in the second and third ones it is rejected (on the basis of consideration of, respectively, the speaker’s abilities and the speaker’s preferences). I show that in each case what the addressee infers about the actual state of the world does not coincide (at least not entirely) with the communicated content of the utterance.

While Henry is cooking in the kitchen, Jane utters (1a). Let us assume that what makes Jane’s utterance relevant is that it implies that Henry should go out and get the dessert. This is something they previously agreed on, so Henry is likely to interpret (1a) as communicating that some of the guests have arrived (where the meaning of ‘some’ is constructed as having a vague cardinality compatible with any number of guests) and (weakly) implicating that Henry should go out and get the dessert.⁶ Henry has no reason to doubt Jane’s trustworthiness and he believes what she communicates to him. Assume, though, that Henry has also planned to buy a bottle of champagne at the local store on the way to the bakery as a surprise for Jane and all the other guests. Jane’s utterance, (1a), will thus license the further inference that he should go and buy the

⁴ For a more detailed discussion about the relationship between comprehension and acceptance see Sperber et al. (2010).

⁵ Sperber et al. (2010) suggested that humans have evolved a suite of cognitive mechanisms which assess the quality of communicated information and the reliability of the individual who dispenses it. The goal of these ‘epistemic vigilance’ mechanisms is the detection of misinformation.

⁶ In fact, Sperber and Noveck (2007) assume that this scalar inference arises at the level of the explicit content of the utterance, as modulation of the meaning of ‘some’ (rather than as an implicature). However, this does not affect the point I am making here. I will discuss this point and other details of the relevance-theoretic account of scalar inferences in Section 4.

champagne before driving to the bakery. Of course, Henry would not take this inference to be intended (Jane is not even aware of the surprise he is organising). It should be clear, then, that the inferences licensed by Jane's utterance go beyond those that Jane may intend to communicate. More generally, the addressee may form a wide range of beliefs about the world through an act of communication, only some of which the speaker may have intended him to form.

Now, consider a different scenario. Henry and Jane have had a fight earlier that day and are not on good terms. After welcoming all the guests, Jane utters (1a) to him. She wants to trick Henry into thinking that not all the guests have arrived, so he will feel under pressure in the kitchen when he later realises that everyone is waiting for the main course which is not ready yet. Suppose that Henry suspects that Jane is being mean and uncooperative towards him. As a consequence, he does not accept her communicated implication *Not all the guest have arrived*. On the contrary, because of his suspicion, he may end up thinking that all the guests have actually arrived. This suggests that not only may the addressee have reasons to doubt the truth of the communicated content (and, as a consequence, fail to accept it as true), but that he may also infer that a state of affairs, which is incompatible with the one described by the communicated content, is likely to hold, on the basis of those very same reasons.

3.2 A methodological note

Let us look now at the implications that this distinction between what is communicated by the speaker and what is accepted/believed by the hearer has for the experimental studies carried out by Bonnefon et al. (2009). I suggest that Experiment 1 and Experiment 2 provide evidence only in support of the claim that, in face-threatening contexts, the addressee is more likely to believe that the possibility that all *X-ed* holds when told that 'Some *X-ed*'. In Experiment 1, participants were more likely to answer "Yes" when asked if it was possible that everyone *X-ed* in a face-threatening context (e.g. "Some people hated your poem") than in a face-boosting context (e.g. "Some people loved your poem"). Experiment 2 provided the same kind of evidence, suggesting that participants judged it as more likely that the speaker could have used the word 'some' while knowing that *all* in a face-threatening context than in a face-boosting-context. The results of these experiments clearly indicate that addressees are more reluctant to rule out the possibility that *all X-ed*, when told that 'Some *X-ed*' in a face-threatening context.

This, however, is not enough to support the stronger claim that they do so because, in face-threatening contexts, speakers do not communicatively intend and addressees do not derive the scalar inference *Not all X-ed*. The same pattern of results could be explained by assuming that while addressees do, in fact, derive the scalar inference *Not all X-ed* as part of the (face-saving) communicated content, they do not accept it as true. The addressee may have reasons to doubt the truth of what the speaker communicates, for example because he thinks that she is trying to be nice and polite rather than strictly honest (as Experiment 3 suggests). If this is the case, he may fail to accept what she communicates (e.g. *Not all X-ed*), and even infer that an alternative state of affairs (e.g. *All X-ed*) is likely to hold on the basis of the same considerations. Interestingly, this explanation receives some support from the results of Experiment 3. These results reveal that, in the face-threatening contexts at issue, the addressee thinks that the speaker is trying to be nice and polite rather than strictly honest. The addressee

may reason that the communicator does not want to hurt his feelings by telling him the whole truth and consequently fail to accept what she communicates as true.⁷

The same point applies to the results of Feeney and Bonnefon (2012). Their participants were presented with a scenario containing an utterance of the form ‘A or B’. They were subsequently asked whether, in their opinion, the scenario ruled out the possibility that A *and* B. The following is a sample scenario for the face-threatening condition (in which A and B are both undesirable events):

- 6 Imagine that Clare is a children’s author. Over lunch her publisher tells her that the sales of her last book have been so poor that she will receive decreased royalties or she will be denied an upfront payment for her next book. In your opinion, does this rule out the possibility that Clare will receive decreased royalties and be denied an upfront payment for her next book?

(Feeney & Bonnefon, 2012, p. 185)

The 10-point scale they were asked to use for their ratings was anchored at *Does not rule out the possibility at all* (lower end) and *Completely rules out the possibility* (higher end). Low ratings can be explained either by assuming that participants adopt an inclusive interpretation of the connective ‘or’ (i.e. they do not derive the scalar inference *A or B but not both*) or by assuming that they derive the scalar inference but reject it as unlikely to be true. In other words, lower ratings may correspond to a lower acceptance rate of the derived scalar inference rather than to the absence of the inference from what is taken to have been communicated by the kind/polite speaker.

It may be asked why a speaker who is concerned to avoid complete factual accuracy in order to save the hearer’s face, e.g. by implicitly communicating that *Not all X-ed* when in fact all X-ed should not straightforwardly utter ‘Not all X-ed’.⁸ After all, if the speaker is willing to sacrifice honesty for kindness, why would she avoid saying explicitly to the addressee “Not all the people hated your poem”? I believe that the choice of uttering ‘Some X-ed’ in these circumstances can be explained by appealing to the notion of ‘deniability’ (Lee & Pinker, 2010). In a context in which the speaker knows that all of the people hated the addressee’s poem, she could implicitly communicate that not all of them did by uttering “Some people hated your poem”, while at the same time keeping open the possibility of denying that she intended to communicate such an implicature. If openly challenged, the speaker could deny that she meant that content, e.g. by suggesting that the semantic (encoded) meaning of ‘some’ (which is compatible with ‘all’) is all that she wanted to communicate. Crucially, speaker denial would require the speaker to pretend that a different conversational context is in place, namely, one in which the assumption that the stronger alternative ‘All X-ed’ would be more relevant to the addressee is not mutually manifest.

More generally, the search for deniability is often motivated by a conflict of interests/goals on the speaker’s side. For instance, in the contexts at issue, while the speaker may want to spare the addressee’s feelings and be polite, she may not want to run the risk of being accused of lying. Implicating is indeed a very important way of merely misleading rather than lying (e.g., Saul, 2012). If openly challenged, the speaker

⁷ This line of explanation relies on the fact that the contexts at issue may plausibly be taken to licence the derivation of the scalar inference. This is true from any theoretical perspective, be it a defaultist account of scalar implicatures or a context-sensitive one (given the relevance of the stronger alternative ‘All X-ed’).

⁸ Thanks to an anonymous reviewer for this journal for raising this question.

can object that what she said was true.⁹

In conclusion, Bonnefon et al.'s (2009) and Feeney and Bonnefon's (2012) results are compatible with two alternative lines of explanation. According to the first, put forth by the authors themselves, they are explained by suggesting that addressees are less likely to derive scalar inferences in face-threatening contexts. According to the second, advanced in this paper, they can be accounted for by assuming that addressees are less likely to accept certain (face-saving) pieces of communicated information as true in a face-threatening context. The crucial distinction between the two is that, while the former assumes that the speaker of a face-threatening utterance of the form 'Some X-ed' is less likely to communicatively intend the scalar inference *Not all X-ed*, the latter maintains that the speaker does intend to communicate this inference but that the addressee may not believe it. It should be clear that these alternative explanations make incompatible claims with regard to the *pragmatic interpretation* of scalar utterances in face-threatening contexts.

3.3 "When some is actually all"

Bonnefon et al.'s (2009) face-management analysis of scalar utterances goes beyond the claim that face-threatening contexts may block the derivation of the scalar inference. They suggest that addressees may interpret a face-threatening utterance like "Some people hated your poem" as *communicating* that *all* people hated his poem. According to this suggestion, when the speaker utters "Some people hated your poem", the addressee may take the proposition *All the people hated your poem* to be communicatively intended. Bonnefon et al. take it as supporting evidence for this analysis that participants (in experiments 1 and 2) are more reluctant to rule out the possibility that *All X-ed* when told that 'Some X-ed' in face-threatening contexts than in face-boosting contexts. In the previous section, I advocated a more cautious analysis of Bonnefon et al.'s results. I questioned whether they provide any conclusive evidence as far as the *interpretation* of the speaker's utterance is concerned. In line with this, I believe that they do not provide any evidence that the addressee takes the proposition *All X-ed* to be *communicatively intended* (i.e. meant) by the speaker. In what follows, I present some arguments that support my rejection of this analysis.

The scalar inference from 'Some X-ed' to *Not all X-ed* can be seen as a special case of a more general phenomenon, that of 'quantity inferences' or 'quantity implicatures' (e.g., Geurts, 2011; Breheny et al., 2013). For example, in a context in which you know that a woman has two sets of objects, forks and spoons, and you know that she is placing them in either one of two boxes, A and B, if a third party tells you "The woman put a spoon in box A and a spoon and a fork in box B", you are likely to infer that the speaker wanted to communicate that the woman put a spoon *and nothing else* in box A (Breheny et al, 2013, p. 425). The inference 'and nothing else' closely resembles the lexically triggered inference *Not all X-ed* from 'Some X-ed' and seems to be derived via the same pattern of reasoning. Now, imagine the following face-threatening context. A and B are working in the same company. During the last year A's sales volume has considerably decreased. The director has informed all the employees that two kinds of cuts will be considered in cases of poor sales performance: retirement benefit and holiday allowance. They are all worried about this. A is speaking with B, the director's secretary.

⁹ For more on the notion of 'saying' which is relevant for the lying-misleading distinction, see Saul (2012).

- 7 A: What did the director decide to do?
B: She will cut your holiday allowance.

Assuming that scalar inferences are a particular case of quantity inferences, Bonnefon et al.'s (2009) analysis of scalar inferences in face-threatening contexts should generalise to (7). The quantity inference *The director will cut your holiday allowance and nothing else* should be less likely to be derived because of the face-threat towards the addressee. Furthermore, B's utterance should be possibly interpreted as communicating the more informative proposition that *The director will cut your holiday allowance and your retirement benefit*. This seems to be an implausible interpretation of B's utterance, but it follows from the face-management analysis provided by Bonnefon and colleagues with regard to scalar inferences. So, it is either that such an analysis generates some counterintuitive predictions (when extended to closely related quantity implicature cases) or that it misses an important generalisation (since the inferential pattern across the cases seems to be the same).

Furthermore, the claim that in face-threatening contexts 'Some X-ed' is interpreted as communicating *All X-ed* is at odds with the empirical findings of Experiment 3. The results of the study show that participants judged the use of 'some', when the addressee knows that *all*, as dishonest and inaccurate both in the face-threatening and in the face-boosting conditions (with the effect being slightly stronger in the face-threatening condition). However, the use of 'some' in the all-condition was considered nice and considerate only in the face-threatening condition. The point worth noting here is that low honesty ratings in the face-threatening condition are arguably incompatible with the claim under discussion. The reason is that if the addressee attributes to the speaker the intention to communicate *All X-ed* in a situation in which it is indeed the case that all X-ed, we wouldn't expect him to judge the communicative act as dishonest. We would expect the addressee to judge the scalar utterance as honest in the face-threatening context but not in the face-boosting one.

To sum up, the claim that 'Some X-ed' is interpreted as *All X-ed* when the context is face-threatening faces two independent challenges: first, it fails to fit within a unified account of quantity implicatures; second, it is not supported by the empirical evidence just surveyed.

4. Relevance Theory and 'scalar inferences'

In their general discussion, Bonnefon et al. (2009) examine whether Relevance Theory (Sperber & Wilson, 1986/1995) is compatible with their face-management analysis of the empirical data. They claim that "the only way [...] for Relevance Theory to account for our results is to make an ad hoc assumption about processing effort" (Bonnefon et al., 2009, p. 257). In order to explain their results, Bonnefon et al. suggest, Relevance Theory would need to assume that interpreting an utterance of 'Some X-ed' as communicating that *All X-ed* requires less processing effort in a face-threatening context than in other contexts. Specifically, in a face-threatening context, the interpretation *All X-ed* would be less costly than the interpretation *Some (but not all) X-ed*, and, because of this, it would be the first interpretation to satisfy the hearer's expectations of optimal relevance and so be attributed to the speaker.

I have already suggested reasons to doubt that hearers interpret scalar utterances in face-threatening contexts as communicating *All X-ed*. In what follows, I argue that

Relevance Theory would not predict such an interpretation but rather the alternative analysis of Bonnefon and colleagues' data which I am advocating in this paper. Before doing this, however, I present the relevance-theoretic account of scalar inferences and discuss the contextual premises that are taken to licence the derivation of scalar implicatures.

4.1 Lexical modulation and scalar expressions

Relevance Theory has long offered a 'deflationary' account of scalar inferences (Sperber & Wilson, 1986/1995; Carston, 1998; Noveck & Sperber, 2007). That is, scalar inferences do not comprise a 'natural kind': not only are they not confined to utterances containing expressions on fixed entailment scales like 'some'/'all' (see Hirschberg (1985) for an account of ad hoc scales), but they are not even inferences of the same type. In particular, they might or might not be part of the communicated content of the utterance, and, when they are, they might arise either at the level of what is explicitly communicated, or as an implicature (i.e. part of the implicit content of the utterance).

According to Relevance Theory, most of the examples of 'scalar inferences' discussed in the literature arise at the level of the explicit content of the utterance (i.e. its explicatures, in relevance-theoretic terms). Specifically, they are conceived of as the result of 'lexical modulation', a pragmatic process "which applies spontaneously, automatically and unconsciously to fine-tune the interpretation of virtually every word" (Wilson, 2004, p. 343). As a result, the concept communicated by the use of a certain word, which emerges as a by-product of the search for relevance, can be significantly different from its lexically encoded meaning. The lexically encoded concept SOME does not include an upper bound (i.e. it denotes any subset containing n elements, where n is at least 2 and at most the total number of elements).¹⁰ However as discussed by Noveck and Sperber (2007), it can be used to communicate a contextually modulated, more specific concept SOME*, whose denotation differs from that of SOME. For instance, it can be used to convey an upper-bounded interpretation of 'some' that is, AT LEAST TWO AND FEWER THAN ALL. While this is certainly a common narrowing down of the encoded meaning of 'some', it is not the only lexical modulation available and it does not carry any special status. Following Noveck and Sperber (2007, p. 192), consider again the following scenario. Jane and Henry are hosting a dinner party. They have agreed that Henry will go out and get the dessert from the pastry shop as soon as the guests start arriving. The doorbell rings and Jane shouts at him from the living room:

- 1 a. Some of the guests have arrived.

Henry does not know how many guests have arrived and whether Jane has already opened the door. Significantly, what makes Jane's utterance relevant is that it implies that Henry should go and get the dessert. This implication is warranted by any construal of 'some'. Thus, Henry's interpretation of (1a) is likely to be compatible with any number of guests, even just one. In this context, the lexically encoded concept SOME works as a clue to derive an *ad hoc* concept SOME**, whose denotation is arguably broader than that of the encoded SOME.

¹⁰ Other authors, including Feeney and Bonnefon (2012), suggested that the semantic meaning of 'some' corresponds to AT LEAST 1 AND POSSIBLY ALL.

It is worth noticing, though, that lexical modulation does not exhaust the phenomenon of ‘scalar inferences’. In some, limited, cases, the ‘scalar inference’ occurs both at the explicit and at the implicit level of communication:

[..] the class of cases described in the literature as scalar inferences is characterized by an enrichment at the level of the explicature (where, for instance, *some* is interpreted in a way that excludes *all*) and only in a small class of these is the exclusion of the more informative concept not just entailed but also implicated. (Noveck & Sperber, 2007, p. 193)

This means that while the narrowing down of the meaning of ‘some’ to SOME* is pretty common, the implication that [*For all the speaker knows*] *not all X-ed* is not always communicated by the speaker (even if it can be derived by the hearer). The crucial distinction to draw here is the one between implication and implicature, which I illustrate with the help of an example (Noveck & Sperber, 2007, p. 192). Consider again (1a), uttered in the following context. While Henry and Jane are waiting for their guests to arrive, Henry is cooking alone. Jane comes in and utters (1a).

- 1 a. Some of the guests have arrived.
- b. Not all of the guests have arrived.

Henry may consider that he should go and greet the guests and bring them some drinks. The fact that Jane’s utterance satisfies the hearer’s expectations of relevance without bringing to mind consequences more typical of the arrival of *all* the guests (e.g. that he should add the pasta to the boiling water), allows Henry to construct the meaning of ‘some’ with a vague cardinality above one and below all. In this context, the explicit content of (1a) implies (in fact entails) (1b) but this may not be actively entertained by Henry (let alone taken to be part of the communicated content of the utterance). Suppose, now, that Henry is wondering whether all the guests have arrived. Given Jane’s utterance, (1a), he may well infer that not all the guests have arrived, i.e. (1b). However, it is only in a context in which there is a mutually manifest reason for the interlocutors to wonder about whether all the guests have arrived, that the denial of this stronger alternative will be implicated. For instance, if Jane has uttered (1a) in response to the question “Have all the guests arrived?”, Henry will take (1a) to implicate that [*For all that Jane knows*] *Not all of the guests have arrived*.

As the examples above suggest, there is a distinction to be drawn among ‘entailment’, ‘inferred entailment’, and ‘communicated entailment’ (or ‘implicated entailment’). While the first notion pertains to a semantic theory, the others arise within a cognitive/pragmatic theory (Carston, 2002, pp. 112-113). Crucially, an entailment whose propositional content is actually entertained (i.e. inferred) by the addressee is not always communicated as an implicature. This occurs only when the utterance achieves relevance by providing an answer to an explicit or tacit question as to whether the stronger alternative (e.g. all the guests have arrived) is true. According to Relevance Theory, scalar inferences are “not scalar, [and] not necessarily implicatures” (Noveck & Sperber, 2007).

4.2 ‘don’t know’ vs. ‘don’t want to say’ implicatures

The derivation of genuine scalar implicatures occurs only if the context makes the stronger alternative a relevant possibility. This is in line with the claim that

particularised approaches conceive of scalar inferences as context-sensitive. A scalar implicature is derived only if it is contextually relevant to the addressee, that is, only if the context is an upper-bound context. However, according to Relevance Theory, the irrelevance of the stronger alternative is not the only contextual factor that may prevent the derivation of scalar implicatures. Considerations about the speaker's 'abilities' (e.g. what she does or does not know) and 'preferences' (e.g. what information she is willing to communicate) may equally block the derivation of scalar implicatures.

The role played by considerations of the speaker's epistemic state (i.e. her 'abilities') has been widely recognised beyond Relevance Theory, being referred in the literature as the 'epistemic-step' in the derivation of scalar implicatures (e.g., Sauerland, 2004; Chierchia, Fox & Spector, 2012). When it is manifest enough that the speaker is ignorant about the more informative alternative (e.g., *All X-ed*), the addressee will usually take her utterance to implicate that she does not know whether the more informative alternative holds – thus, the 'don't know' implicature. Much less attention has been devoted to the role played by non-epistemic mental states. In this respect, Relevance Theory represents a significant exception. The relevance-guided interpretation process is said to be constrained by both the speaker's 'abilities' and the speaker's 'preferences'. The latter correspond to a wide range of goals in addition to the fundamental communicative goal (e.g. the desire to withhold some information from the interlocutor). In virtue of this, Relevance Theory is able to offer an integrated account of the different kinds of implicature which may arise from an utterance of the form 'Some of the X...'. Among these are the class of 'don't want to say' implicatures, like the one generated by B's utterance (in an appropriate context) in the following exchange:

- 8 A: Which of your colleagues support the strike?
 B: Some of them do.

(Carston, 1998, p. 271)

If it is mutually manifest that B could have been more specific (e.g. if it is mutually manifest that B knows who among her colleagues supports the strike), then her answer, together with the presumption that her utterance would be the most relevant one compatible with her abilities and preference, will imply that she is *unwilling* to be more specific, that is, that she doesn't want to say which colleagues support the strike.

To sum up, according to Relevance Theory, an utterance containing the scalar expression 'some' does not necessarily give rise to a scalar implicature of the form 'not all...'. The scalar implicature is derived if the context satisfies the following constraints: it is upper-bound, that is, the relevance of the stronger alternative is mutually manifest to the interlocutors, and the speaker is both willing and able to provide the hearer with information about the stronger alternative.

4.3. Relevance Theory and face-threatening contexts

I now turn to Relevance Theory and its predictions regarding the effect of politeness on scalar inferences. In section 4.1, I emphasised that the derivation of genuine scalar implicatures crucially depends on contextual premises as to whether the stronger alternative is entertained as a relevant possibility. In other terms, the scalar implicature '[The speaker believes that] not all X-ed' is derived only if the alternative 'All X-ed' is entertained as a relevant possibility in the context at issue. In the contexts set up by

Bonnefon et al. (2009) in the three experiments discussed above, we may assume that it is mutually manifest among the interlocutors that the alternative ‘All X-ed’ (e.g. all members hated your poem) would be more relevant to the addressee (who, in all the three experiments, coincides with the participant himself/herself). According to Relevance Theory, this leads to three different possibilities: (a) the derivation of the scalar implicature ‘[The speaker believes that] not all X-ed’, (b) the derivation of the ‘don’t know’ implicature (i.e. ‘The speaker does not know whether all X-ed’), and (c) the derivation of the ‘don’t want to say’ implicature (i.e. ‘The speaker is reluctant to say that all X-ed’). The interpretative route is chosen among (a)-(c) on the basis of context-specific considerations about the speaker’s ‘abilities’ and ‘preferences’.

Consider again the following scenario (Experiment 1):

- 4 Imagine that you have joined a poetry club, which consists of five members in addition to you. Each week one member writes a poem, and the five other members discuss the poem in the absence of its author. This week, it is your turn to write a poem and to let the others discuss it. After the discussion, one fellow member confides to you that “Some people hated your poem”.

(Bonnefon et al., 2009, p. 251)

According to Relevance Theory, if it is mutually manifest that the speaker has complete knowledge, the addressee would either infer that the speaker is unwilling to be more specific or that she believes that the stronger alternative ‘All of the people hated your poem’ is false. If either of these possibilities is manifest and relevant, it will be implicated. In the former case, the hearer would derive the ‘don’t want to say’ implicature (i.e. *Denise is reluctant to say that all the people hated the poem*); in the latter he would derive the ‘scalar implicature’ (i.e. *[Denise believes that] Not all of the people hated the poem*). The possibility that the speaker is unwilling to be more specific could be manifest to the hearer in circumstances in which the speaker is clearly trying to be elusive and vague.¹¹ So, in a face-threatening context like (4), a speaker attempting to save the face of the hearer in such a way may increase the manifestness of the possibility that she is unwilling to be more specific. As a consequence, the hearer may derive the ‘don’t want to say’ implicature. This is, however, a somewhat peculiar scenario since it requires the speaker’s politeness/face-saving concerns to be mutually manifest to the speaker and the hearer.

In all other circumstances, it seems plausible to assume that the possibility that the speaker believes that not all X-ed would be manifest enough to make it mutually manifest that the speaker intended it to be manifest, since it increases the relevance of her utterance and is compatible with her manifest preferences. As suggested in section 3, however, the addressee may end up rejecting this communicated content on the basis of considerations about the speaker’s ‘preferences’. He may fail to accept the scalar inference as true because he believes that Denise’s preference for politeness rather than strict honesty underlies her communicating to him that not all of the people hated his poem. As a consequence, he may not rule out the possibility, incompatible with the scalar inference, that all of the people hated his speech (and that Denise believes this to be so). In this case, considerations about the speaker politeness ‘preferences’ do not

¹¹ For instance, the speaker may rely on prosodic clues such as the use of a B-accent on ‘some’ (Jackendoff, 1972), which usually indicates that the speaker is unwilling to utter (at least) one alternative (for reasons that need to be further contextually specified).

affect the interpretation of the utterance, but rather the assessment of its believability.¹²

In conclusion, when the speaker is assumed to be in a position to know whether the more relevant alternative ‘All X-ed’ holds, the possible effects of politeness on the derivation of scalar inferences *Not all X-ed* are diverse.¹³ If it is mutually manifest that the speaker has an opinion as to whether ‘All X-ed’ is true (whose content is not manifest to the addressee), the addressee may derive either the ‘don’t want to say’ implicature, or the scalar implicature [*The speaker believes that*] *not all people hated my poem*. In the former case, which is presumably less common, politeness triggers the derivation of the ‘don’t want to say’ implicature; in the latter, it puts under discussion the believability of the communicated scalar inference.

Summing up: while Bonnefon and colleagues present a very interesting set of data, whose explanation needs to be integrated within theoretical pragmatics frameworks, the contexts adopted in their studies do not provide the means for discriminating Relevance Theory’s different predictions on the effect of politeness on the derivation of scalar implicatures. Furthermore, Relevance Theory does not predict that face-threatening utterances of the form ‘Some X-ed’ would be interpreted as conveying *All X-ed* in any context. However, this is the only interpretation that Bonnefon et al. (2009) try to account for in light of the relevance-theoretic account. It comes as no surprise, then, that the result falls short of theoretical adequacy.

Recall that one of the aims of Bonnefon et al. (2009) is to broaden the empirical scope of the debate between ‘defaultist’ and ‘context-sensitive’ accounts of scalar inferences by identifying new contexts which could tell apart the predictions of the two accounts. With this in mind, it is interesting to notice that Relevance Theory seems to predict that in the face-threatening contexts set up by Bonnefon et al. the scalar utterance always gives rise to an implicature (be it a ‘don’t want to say’ implicature or a scalar implicature). If this is the case, the contextual manipulation investigated here may not provide an appropriate means to test alternative theories of scalar inferences (‘default’ accounts vs. ‘context-sensitive’ accounts). The problem here is that, even if it may (occasionally) make the scalar inference inappropriate, the expected speed of interpretation of ‘some’ may not vary across the two models. While default accounts predict that the time taken to derive the interpretation would increase in virtue of the extra cognitive effort involved in cancelling the inappropriate scalar inference (which has arisen automatically), relevance theorists predict that the time taken would (equally?) increase because of the cognitive effort involved in deriving the ‘don’t want to say’ implicature.¹⁴

5. Honesty

As well as exploring the effect of politeness on the derivation of scalar inferences, Feeney and Bonnefon (2012) investigate the role of (self-rated) honesty. Their

¹² For a detailed discussion of the role of speakers’ ‘abilities’ and ‘preferences’ in the pragmatic process of interpretation and in the assessment of the believability of the communicated content, see Mazzarella (2013).

¹³ If it is mutually manifest that the speaker does not have an opinion as to whether ‘All X-ed’ holds, the addressee is likely to derive the ‘don’t know’ implicature (e.g. *Denise does not know whether all of the people hated your speech*). In this case, considerations about politeness do not enter the picture.

¹⁴ More generally, this problem seems to extend to context-sensitive accounts other than Relevance Theory. It is usually assumed that upper bound contexts in which the derivation of the scalar implicature is blocked give rise to implicatures of other sorts (e.g. non-epistemic quantity implicatures, see Geurts (2011, pp. 35-36)).

hypothesis is that more honest people may be more likely to derive scalar inferences (e.g. to interpret the connective *or* in ‘A or B’ as *A or B but not both*). As previously discussed, there are reasons to doubt that the experiments at issue address the *interpretation* of scalar utterances (rather than the believability of the inference they may warrant). I present Feeney and Bonnefon’s analysis of the results and suggest an alternative explanation of the data.

Let us first investigate the theoretical motivations behind the hypothesis that honesty and scalar inferences positively correlate. They revolve on two claims: (i) because of ‘social projection’ (Robbins & Krueger, 2005), people who consider themselves honest are more likely to expect other speakers to manifest the same honesty¹⁵, (ii) scalar inferences are facilitated by an assumption of honesty. Feeney and Bonnefon suggest that honest speakers are more cooperative speakers, that is, they try to be as informative as possible given their epistemic state. When using an expression that is relatively low on a salient scale, such as *or*, they can be expected not to be in the epistemic position to make a stronger claim (e.g. ‘A and B’). Thus, hearers can confidently derive the scalar inference *A or B but not both*. It follows that since more honest people may tend to attribute a higher degree of honesty to their interlocutors (see (i)), they may be more likely to derive scalar inferences when their interlocutors use a scalar expression.

The experiments seem to support this hypothesis. Regardless of context (be it face-threatening or face-boosting), more honest people are more likely to rule out the possibility that A and B, given an utterance of ‘A or B’. In other terms, while both Face and Honesty had a main effect, no interaction was detected. In what follows, I put forth an alternative explanation of the main effect of Honesty. Once again, the principal aim of this suggestion is to show that the experimental evidence provided does not warrant Feeney and Bonnefon’s explanation over this alternative.

My argument focuses on the relationship between honesty and trust (see Sperber et al. (2010)). Honest people are usually trustworthy: they are willing to share information that they possess and regard as true. This is the reason why they are more likely to be granted epistemic trust. Epistemic trust can be defined as the willingness to believe the communicator and accept her claims as true. It follows that hearers are more likely to accept a piece of communicated information when this comes from an interlocutor they judge to be honest.

As suggested in section 3, the experimental paradigm adopted by Bonnefon et al. (2009) and Feeney and Bonnefon (2012) does not allow us to draw firm conclusions about the selected *interpretation* of scalar expressions (in face-threatening and face-boosting utterances). Given a sentence of the form ‘A or B’, participants are asked whether the possibility that *A and B* is ruled out. While the authors suggest that lower ratings on the scale ‘*Does not rule out the possibility at all* (lower end) - *Completely rules out the possibility* (higher end)’ correspond to an inclusive interpretation (i.e. to the absence of the scalar inference), I suggested that they may depend on a lower acceptance rate of the scalar inference (i.e. the exclusive interpretation). The scalar inference would be derived and attributed to the speaker (as part of her intended meaning) but not accepted as true.

If this is the case, however, the correlation between honesty and higher ratings comes as no surprise. If honest people are more likely to accept a piece of communicated information because they are more likely to grant epistemic trust to their interlocutor (because of social projection), they will tend to rule out the possibility that

¹⁵ In what follows, I assume the validity of (i). For more details, see Robbins & Krueger (2005).

A and B when the speaker has uttered ‘A or B’. This would not depend on the fact that they are more likely to derive the scalar inference than dishonest people, but rather that they are more likely to accept it as true when it is communicated.

This explanation is compatible with the pattern of results and with the absence of interaction between Face and Honesty. Honest people are generally more likely to accept a piece of communicated information (regardless of context) but they are not “any less sensitive to the demands of politeness” (Feeney & Bonnefon, 2012, p. 187) than less honest individuals.

6. Conclusions

The literature in experimental pragmatics has fruitfully focused on scalar inferences and investigated how contextual factors modulate their derivation. Among these, the effects of politeness and face-threatening contexts have been recently explored and politeness contexts have been claimed to reduce the derivation of scalar inferences (similarly to lower-bound contexts). To my knowledge, this conclusion has been accepted as uncontroversial (e.g., Cummins, 2012).

In this paper, I have raised methodological and theoretical issues with regard to the studies on politeness and scalar inferences proposed by Bonnefon and colleagues (Bonnefon et al, 2009; Feeney & Bonnefon, 2012). The experimental evidence provided so far falls short of demonstrating that scalar inferences are less likely to be derived in face-threatening contexts. In fact, it is compatible with an alternative explanation: politeness may affect the believability (rather than the derivation) of the scalar inferences. In face-threatening contexts, addressees may be less likely to accept the scalar inference as true because they recognise that the speaker is being polite. Furthermore, I suggested that Relevance Theory makes predictions that are compatible with this line of explanation (but not with the one proposed by Bonnefon et al. 2009).

The crucial distinction to be drawn here is the one between the speaker’s intended meaning (i.e. the communicated content of the utterance) and the beliefs about the world that the hearer adopts on the basis of that communicated content. Significantly, the hearer may attribute an interpretation to the speaker without accepting it as true. Even more, he may end up forming a belief that is incompatible with the content communicated by his interlocutor. When told that ‘some’ of the people hated his poem, the hearer may end up believing that *all* of them did, even if he attributes to the speaker the intention to convey that *some but not all* the people hated the poem.

The very same distinction can be used to undermine the experimental evidence proposed to support the hypothesis that (self-rated) honesty positively correlates with the derivation of scalar inferences. Far from showing the effect of individual differences in the *interpretation* of scalar expressions, Feeney and Bonnefon’s results are (at least) compatible with a very different explanation: more honest people may be more likely to accept the scalar inference as true because of their tendency to grant epistemic trust to the interlocutor.

Interestingly, this line of explanation opens up a new way of thinking about certain cases of understatement.¹⁶ Understatement (or meiosis) is a rhetorical trope which is typically analysed as the opposite of hyperbole. We think of understatement as a way to communicate something stronger in an indirect way (e.g., of a man known

¹⁶ Thanks to an anonymous reviewer for this journal for pointing this out to me.

to have broken up all the furniture, one says: “He was a little intoxicated”, Grice 1975/1989, p. 34). However, let us focus on the subset of ‘polite understatments’, that is, understatments which might serve a face-saving function. Consider the following example.

9 Not everybody thought that you were sober.

The usual view would be that, by uttering (9), the speaker is implicating the stronger proposition that *Nobody thought you were sober*.

In line with the analysis of face-threatening scalar utterances proposed above, there is an alternative account of examples like this. While the semantic (encoded) meaning of ‘not everybody’ is compatible with ‘nobody’, it is possible to argue that (9) has a (non-prototypical) scalar implicature *It is not the case that nobody thought that you were sober*, which is truth-conditionally equivalent to *Some thought that you were sober*. That is, polite understatments do have a ‘polite’ implicature, but this may not be accepted as true by the addressee. Specifically, the ‘polite’ implicature would not be taken to be sincere and, as a consequence, the addressee would not accept its content as true and would further infer that it is likely to be the case that *nobody* thought that he was sober. Interestingly, this line of reasoning might be anticipated, even intended, by the speaker, without the inference to *Nobody thought that you were sober* falling under her communicative intention (in the sense of ‘ostensive-inferential’ communication proposed by Sperber and Wilson (1986/1995)).

In conclusion, while Bonnefon and colleagues’ studies open a new interesting direction for empirical research on scalar inferences, the question of the effect of politeness on the derivation of scalar inferences is far from being settled. Future research in experimental pragmatics should aim at capturing the distinction between comprehension and acceptance, which theoretical pragmatics has long embraced (e.g., Austin, 1962; Grice, 1957; Sperber & Wilson, 1986/1995) and recently enriched with new insights (Sperber et al., 2010).

Acknowledgments

I am grateful to Robyn Carston for her valuable guidance and insightful discussions about the content of this paper. Thanks also to two anonymous reviewers for their constructive comments and suggestions. This work is supported by the Leverhulme Trust (Ref: F/07 134/DP).

References

- Austin, John L. (1962). *How to do things with words*. Cambridge, MA: Harvard University Press.
- Bonnefon, Jean-François, Feeney, Aidan, & Villejoubert, Gaëlle. (2009). When some is actually all: Scalar inferences in face-threatening contexts. *Cognition*, 112, 249-258.
- Brehyeny, Richard, Katsos, Napoleon, & Williams, John. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, 100, 434-463.

- Breheeny, Richard, Ferguson, Heather J., & Katsos, Napoleon. (2013). Taking the epistemic step: Towards a model of on-line access to conversational implicatures. *Cognition*, 126, 423-440.
- Brown, Penelope, & Levinson, Stephen C. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press (Original work published 1978).
- Carston, Robyn. (1990). Quantity maxims and generalised implicature. *UCL Working Papers in Linguistics*, 2, 1-31. Reprinted in *Lingua*, 96, 213-244.
- Carston, Robyn. (1998). Informativeness, relevance and scalar implicature. In Carston, R., Uchida, S. (Eds.), *Relevance Theory: Applications and Implications* (pp.179-236). Amsterdam: John Benjamins.
- Carston, Robyn. (2002). *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.
- Chierchia, Gennaro. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and beyond: The cartography of syntactic structures* (Vol. 3, pp. 39-103). Oxford: Oxford University Press.
- Chierchia, Gennaro, Fox, Danny, & Spector, Benjamin. (2012). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In P. Portner, C. Maienborn, & K. von Stechow (Eds.), *Handbook of semantics* (vol. 3, pp. 2297–2331). Mouton de Gruyter.
- Cummins, Chris. (2012). Modelling implicatures from modified numerals. *Lingua*, 132, 103-114.
- Feeney, Aidan, & Bonnefon, Jean-François. (2012). Politeness and honesty contribute additively to the interpretation of scalar expressions. *Journal of Language and Social Psychology*, 32(2), 181-190.
- Geurts, Bart. (2011). *Quantity Implicatures*. Cambridge: Cambridge University Press.
- Grice, Herbert Paul. (1957). Meaning. *Philosophical Review*, 66, 377-388.
- Hirschberg, Julia. (1985). *A theory of scalar implicature*. Ph.D. thesis. University of Pennsylvania.
- Horn, Laurence. (1972). *On the semantic properties of logical operators in English*. PhD dissertation, University of California, LA.
- Horn, Laurence. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Shiffrin (Ed.), *Meaning, form, and use in context* (pp. 11-42). Washington, DC: Georgetown University Press.
- Jackendoff, Ray. (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.
- Katsos, Napoleon. (2008). The semantics/pragmatics interface from an experimental perspective: the case of scalar implicature, *Synthese*, 165, 358-401.
- Katsos, Napoleon, Breheeny, Richard, & Williams, John. (2005). The interaction of structural and contextual constraints during the on-line generation of scalar inferences. In B. Bara, L. Barsalou and M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 1108-13). Mahwah, NJ: Erlbaum.
- Lee, James J., & Steven Pinker (2010). Rationales for indirect speech: The theory of the strategic speaker. *Psychological Review*, 117(3), 785-807.
- Levinson, Stephen C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.
- Mazzarella, Diana. (2013). ‘Optimal relevance’ as a pragmatic criterion: The role of epistemic vigilance. *UCL Working Papers in Linguistics*, 25, 20-45.

- Noveck, Ira, & Sperber, Dan. (2007). The why and how of experimental pragmatics: the case of 'scalar inference'. In N. Burton-Roberts (Ed.), *Advances in Pragmatics*. (pp. 184-212). Basingstoke: Palgrave.
- Robbins, Jordan M., & Krueger, Joachim I. (2005). Social projection to ingroups and outgroups: A review and meta-analysis. *Personality and Social Psychology Review*, 9, 32-47.
- Sauerland, Uli. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy*, 27, 367-391.
- Saul, Jennifer M. (2012). Lying, misleading and what is said: An exploration in philosophy of language and in ethics. Oxford: Oxford University Press.
- Sperber, Dan, & Wilson, Deirdre. (1987/1995). *Relevance: Communication and cognition*. Oxford: Blackwell.
- Sperber, Dan, Clément, Fabrice, Heintz, Christophe, Mascaro, Olivier, Mercier, Hugo, Origi, Gloria, & Wilson, Deirdre. (2010). Epistemic vigilance. *Mind & Language*, 24(4), 359-393.
- Wilson, Deirdre. (2004). Relevance and lexical pragmatics. *UCL Working Papers in Linguistics*, 16, 343-360.
- Wilson, Deirdre, & Carston, Robyn. (2007). A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts. In N. Burton-Roberts (Ed.), *Advances in Pragmatics*. (pp. 230-260) Basingstoke: Palgrave.
- Wilson, Deirdre, & Sperber, Dan. (2004). Relevance Theory. In L. Horn, G. Ward (Eds.) *Handbook of Pragmatics*. (pp. 607-632). Oxford: Blackwell.