TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science (2019) 1–25

ISSN:1756-8765 online
DOI: 10.1111/tops.12456

This article is part of the topic "Lying," Hans van Ditmarsch, Petra Hendriks, and Rineke Verbrugge (Topic Editors). For a full listing of topic papers, see http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1756-8765/earlyview

Strategies of Deception: Under-Informativity, Uninformativity, and Lies—Misleading With Different Kinds of Implicature

Michael Franke,[a] Giulio Dulcinati,[b] Nausicaa Pouscoulous[b]

[a]*Institute of Cognitive Science, University of Osnabrück*
[b]*Division of Psychology and Language Sciences, University College London*

Received 16 October 2017; received in revised form 17 August 2019; accepted 17 August 2019

Abstract

Conversation is often cast as a cooperative effort, and some aspects of it, such as implicatures, have been claimed to depend on an assumption of cooperation (Grice, 1989). But any systematic class of inference derived from assumptions of cooperation, such as implicatures, could also be, on occasion, used to deceive listeners strategically. Here, we explore the extent to which speakers might choose different kinds of implicature triggers in an uncooperative game of communication. Concretely, we present a study in the form of a cooperative or competitive signaling game where communicators can exploit three kinds of implicatures: exact reading of numeral expressions, scalar implicatures linked to the quantifier *some* and *ad hoc* scalar implicatures. We compare how these implicatures are used depending on whether the participants' co-player is cooperative, a strategic opponent, or a non-strategic opponent. We find that when the strategy of the co-player is clear to the participants, the three types of implicatures are used to exploit the co-player's interpretation strategy. Indeed, participants use numeral implicatures as reliably as truth conditional content in all three conditions, while scalar quantifiers and *ad hoc* implicature elicit different strategies. We interpret these findings as evidence that speakers expect their interlocutors to infer implicatures from their utterances even in contexts where they know that they will be perceived as uncooperative.

Keywords: Lying; Implicature; Signaling game; Cooperation; Grice

Correspondence should be sent to Michael Franke, Institute of Cognitive Science, University of Osnabrück, Wachsbleiche 27, 49069 Osnabrück, Germany. E-mail: mchfranke@gmail.com

1. Introduction

Grice (1989) famously presents conversation as a cooperative activity in which participants abide by a *cooperative principle*, which binds them to make appropriate contributions to the conversation. From this principle follow more specific maxims such as the first maxim of quantity: "Make your contribution as informative as is required" (Grice, 1989, p. 45). Speakers can exploit the maxims in order to communicate implicit propositions (implicatures) of various types. For example, the speaker can violate the first maxim of quantity to communicate a quantity implicature. If I tell you that *I used some of your new shampoo* in a context where it would be relevant and more informative to know whether *I used all of your shampoo*, you may infer that the reason why I am violating the first maxim of quantity is that the more informative statement is not true and therefore infer the implicature that *I did not use all of your shampoo*.

The last two decades witnessed a wave of experimental investigation of how different types of quantity implicatures are processed and interpreted; and in harmony with Grice's account, these investigations have focused on situations where the cooperation and honesty of the speaker is taken for granted. However, conversation also takes place in non-cooperative or competitive situations, where the speaker may be deceptive or uninformative. Politicians are often good examples of unhelpful interlocutors. For instance, consider this evasive answer that Theresa May gave in 2016 when asked whether the United Kingdom should have access to the EU single market after Brexit: "What I want to see is the best possible deal for the United Kingdom in trade in goods and services" (Bull, 2016). The use and comprehension of implicatures in non-cooperative settings is a vastly understudied topic. The very few existing comprehension studies on this topic suggest that listeners faced with an uncooperative speaker tend to infer less implicatures than if they are faced with a cooperative speaker (Dulcinati, 2018; Dulcinati & Pouscoulous, 2017; Pryslopska, 2013). Even less work has systematically investigated implicatures in uncooperative contexts from a speaker's production perspective. We are aware of only one study which did this (Mol, Verbrugge, & Hendriks, 2005; Verbrugge & Mol, 2008). It had participants play a variant of Mastermind. Players took turns to guess the co-player's secret sequence of four colors. After each guess, players reported to their co-players the number of matching colors in correct and incorrect positions. In doing so, the reporting player had to choose descriptions from a predefined list. Reporting players had an incentive to be uninformative or misleading, because they would win if their opponent failed to guess their color sequence correctly after a fixed number of moves. The authors found that some speakers, but not all, tended to select under-informative quantifier expressions.

The study presented in this paper is similar in spirit but goes beyond this work in that it explicitly compares production of different kinds of quantity implicatures and it also provides a baseline cooperative condition for comparison. We believe that it is particularly interesting and timely to fill this gap in the literature, first because this method allows us to compare how speakers use explicit and implicit communication strategically, and second because it may offer a new perspective on the differences between well-studied types of quantity implicatures.

Non-cooperative speakers may differ from cooperative ones in that they may be more likely to deceive or to be uninformative. Although Grice (1989) presents conversation as a cooperative effort, he contemplates both the possibility that speakers may be uninformative by opting out of the cooperative principle or of a maxim in an overt way, for example, by saying "I can't tell you that," and the possibility that they may be deceitful by *covertly* violating a maxim. The paramount example of covert violations of maxims is lying, where the liar covertly violates the first maxim of quality (i.e., "Do not say what you believe to be false"; Grice, 1989, p. 46) and intends the audience to remain unaware of the violation. Besides lying, the realm of verbal deception includes falsely implicating. While to lie, at least according to traditional definitions (Isenberg, 1973; Primoratz, 1984), is to *say* something that the speaker believes to be false with the intention to deceive; to falsely implicate is to communicate something believed to be false by means of a *conversational implicature* (Meibauer, 2014). For example, if I said that "*I used some of your new shampoo*" when in fact I believe that I used all of it, I could be falsely implicating that *I did not use all of your new shampoo*. Although there is an ongoing conceptual debate on whether false implicatures should be considered *lies* (Meibauer, 2005, 2014) or not (Dynel, 2011, 2015), we will treat them here as separate for the purposes of experimental design and analysis. We let the data speak about any potential difference between false implicatures and lies. Therefore, the phenomena which we expect to observe with possibly different behavioral signatures in our study are uninformativity or Gricean opting out, lies, and false implicatures.

Lies are part of what is explicitly communicated, while false implicatures are part of the realm of implicit communication (see Carston, 2002, 2009 and Recanati, 2004 for a review of the implicit/explicit divide).[1] This distinction is likely to impact communication in non-cooperative contexts and whether a speaker goes for telling a lie or, merely, conveying a false implicature. We expect that the use of lies and false implicatures might differ in our study because some aspects of implicit communication make it particularly advantageous in non-cooperative contexts. Indeed, implicit communication offers the advantage of *plausible deniability* (Pinker, Nowak, & Lee, 2008): Implicatures are cancellable and therefore speakers can deny having intended to communicate them. I can say, *I used some of your shampoo* and later claim I did not mean that I did not finish it. Such denial is, of course, impossible if I explicitly stated that I did not use all of your shampoo. This feature of implicit communication comes in handy when the speaker wants to communicate something that could incur a penalty, such as proposing a bribe or providing false information. According to Reboul (2017), implicit communication also offers another advantage: Hearers might endorse information communicated implicitly more easily than explicit content for two reasons. First, hearers are more vigilant toward content, the speaker is strongly committed to, and a higher degree of speaker commitment is carried by explicit, not implicit, content (Morency, Oswald, & de Saussure, 2008). Second, hearers are less vigilant toward content that is the fruit of their own inferences, which is the case for implicatures but not for asserted content. All these reasons might push a speaker to prefer false implicatures to explicit lying.

Some previous studies that are similar to the one we are presenting looked at non-verbal deception in the context of signaling games where signalers have to give non-verbal hints (e.g., images, maps) to a receiver player who has to make choices based on the information provided in the hints (Montague, Navarro, Perfors, Warner, & Shafto, 2011; Ransom, Voorspoels, Perfors & Navarro, 2017). Crucially in some cases, the game is competitive and the signaler benefits from the receiver's wrong choices, which provides motivation to deceive. Signalers can give true hints, false hints, uninformative hints, and misleading hints, which like false implicatures consist in conveying a true piece of information which leads the receiver to infer something false. Montague et al. (2011) found that their players preferred giving misleading hints rather than false hints. In their game, the receiver did not know whether the signaler was cooperative or competitive and they could choose to check whether the hints were false and calibrate their trust accordingly, which was an incentive for the signaler not to be caught lying as it would have reputation consequences for the rest of the game. In a similar competitive game, Ransom et al. (2017) gave their participants the options to give to the receiver true, misleading, or uninformative visual hints, but not false hints, and they manipulated the signaler's expectations regarding how suspicious or trustful the receiver would be. Because the receiver did not know whether their signaler was honest or deceitful, the signaler could pretend to be helping the receiver while in fact providing them false or misleading information. They found that when signalers expected a trustful receiver, they were more likely to mislead, whereas when they expected a suspicious receiver they were more likely to be uninformative.

These studies are similar to ours first because we also employed a competitive signaling game and secondly because the types of deception they studied use the same fundamental mechanisms as the kinds of verbal deception we are interested in, which are to cause someone to have a false belief (Mahon, 2007) either by communicating something false (i.e., false hints, lies) or by communicating something true (i.e., misleading hints, false implicatures).

One important complication of studying explicit and implicit cases of verbal deception is that while the studies on non-verbal deception that we mentioned could draw a clear distinction between false and misleading hints, drawing a distinction between lies and false implicature is not straightforward. In two studies where participants were asked to rate a false implicature on a scale that ranged from an utterance being a lie to an utterance not being a lie, the average rating was near the middle of the scale (Coleman & Kay, 1981; Hardin, 2010). In parallel to these results, studies on the explicit–implicit distinction in comprehension found that lay-people are likely to consider implicatures part of *what is said* under some circumstances (Doran, Baker, McNabb, Larson, & Ward, 2009; Doran, Ward, Larson, McNabb, & Baker, 2012; Nicolle & Clark, 1999). Doran et al. (2012) asked participants to judge whether sentences that could give rise to an implicature were true or false in light of a fact contradicting the implicature (e.g., judging whether the sentence *I used some of your shampoo* is true given that I used all of it).

Implicatures arising from cardinal numbers (e.g., *I have three cats* implicating that I do not have four) were incorporated into the truth conditional meaning of the utterance 53% of the time, while for scalar implicatures (triggered by quantifiers such as *some* and *most*) this happened only in 32% of cases. Weissman and Terfourafi (2019) used a similar paradigm to investigate whether false implicatures are considered to be lies, looking at several different types of implicatures. They asked participants to evaluate utterances carrying a false implicature on a 7-point scale ranging from "definitely not a lie" to "definitely a lie." Although most implicatures types were not found to be considered lies, they found like Doran et al. (2012) a difference between numerals and scalar quantifiers: False numeral implicatures were one of the only cases to be clearly considered lies, while their quantifier utterances were rated as neither lies or not lies. These findings all suggest that different types of implicatures differ in how likely they are to be considered part of explicit communication and therefore in whether they are considered to be lies when used deceptively. Different implicature triggers also differ in how easily they give rise to implicatures (van Tiel, van Miltenburg, Zevakhina, & Geurts, 2016). This variability is interesting in and of itself. We wanted to explore it, but we also wanted to ensure that our findings would not be restricted to a single implicature type. We therefore chose to use three different types of implicatures in our study.

We focused on three particularly interesting types of quantity implicatures: implicatures arising from numerals, from the quantifier *some,* and from *ad hoc* constructions, which can all be viewed as types of scalar implicatures (Hirschberg, 1991; Horn, 1972, 1989; van Rooij & Schulz, 2006). For all of them, the implicature arises from the negation of an alternative utterance containing a stronger term on the same semantic scale. In the case of the quantifier *some* and numerals, these scales are lexicalized and the upper-bound readings in 1a and 2a below, for instance, come from the negation of the stronger alternative in 1b and 2b, respectively. Yet, in the case of *ad hoc* scalar implicatures there is no pre-established lexical scale and the scale arises from the context. No lexical scale links Laurel to Hardy, thus in 3a the implicature "I didn't buy a present for Hardy" could only be derived in a context where the pair is particularly salient and buying them both a present would be relevant.

1a. I used *some* of your shampoo.
1b. I used *all* of your shampoo
2a. I have *three* cats.
2b. I have *four* cats.
3a. I bought a present for *Laurel*.
3b. I bought a present for *Laurel and Hardy*.

There is ongoing theoretical controversy about the nature of inferences associated with different putative implicature triggers. Implicatures linked to *ad hoc* scales are undoubtedly *particularized* implicatures in Gricean (1989) terms since they can only be intended in a context that makes the *ad hoc* scale relevant. By contrast, both scalar terms and numerals have given rise to much theoretical controversy about whether their upper-bound interpretation is the result of an entirely context-dependent implicature or whether it is part of their semantic or *default* meaning (Geurts, 2010; Levinson, 2000). Some theorists argue that lexicalized scalar implicatures linked to quantifiers are an output of grammar (Chierchia, 2004; Chierchia, Fox & Spector, 2012), while others maintain they result from a pragmatic inference. In the latter camp, some defend they are *particularized* implicatures (like *ad hoc* scalar implicatures) and are derived when prompted by context (Geurts, 2010; Noveck & Sperber, 2007), while others maintain they are *generalized* implicatures and that they arise systematically unless the context blocks them (Horn, 1989) or even by default (Levinson, 2000). The interpretation of numerals is also hotly debated with some theorists claiming that they have a lower-bound or an "*at least*"-meaning while the *exact* interpretation is supplied in the context via an implicature (e.g., Gadzar, 1979; Horn, 1972; Levinson, 2000) and others claiming that the *exact* interpretation of numerals is not an implicature but part of their truth-conditional meaning (Breheny, 2008; Carston, 1998; Geurts, 2006; Kennedy, 2015). It is therefore unclear whether we should lump lexical quantifier scales with numerals, with *ad hoc* scales, or whether these are three entirely distinct categories of implicatures.

Empirical evidence has not fully settled the debate about differences between implicature triggers either. It provides some fairly clear evidence setting numerals apart from quantifier scalar implicatures. Papafragou and Musolino (2003) and Huang, Spelke, and Snedeker (2013) provide convincing evidence that numerals have an exact truth-conditional interpretation (i.e., three means "exactly three") by showing that pre-school children give exact interpretations of numerals despite finding scalar implicatures notoriously difficult (Noveck, 2001; Papafragou & Musolino, 2003; Pouscoulous, Noveck, Politzer, & Bastide, 22007). In the same vein, Huang and Snedeker (2009) found that while adults process the upper bound meaning of scalar terms such as *some* more slowly than the literal meaning of the quantifier *all*, they process the *exact* meaning of numerals just as quickly; suggesting that the former involves drawing a pragmatic inference while the second does not. The comparison between lexical and *ad hoc* scalar implicatures offers mixed results. Recent experimental evidence suggests that the scalar implicatures linked to quantifiers are derived in the same way as particularized quantity implicatures (see Katsos & Cummins, 2010 for a review). Yet part of the developmental literature suggests that children have more trouble understanding scalar implicatures with quantifiers than *ad hoc* implicatures (Barner, Brooks, & Bale, 2011; Stiller, Goodman, & Frank, 2011), while in contrast Katsos (2009) suggests they calculate both types of implicatures to the same extent. Finally, somewhat in opposition to previous studies, Rees and Bott (2018) found lower rate of implicature readings for *ad hoc* expressions than for the quantifier *some* or numerals in a study using a structural priming paradigm. Therefore, despite intensive theoretical and experimental interest in these quantity implicatures for the past decades, the layout of numerals, quantifiers, and *ad hoc* scales remains unclear. Some empirical findings suggest that they do not always align with each other. When it comes to uncooperative language use and language processing, exact readings of numerals seem to behave more like literal meaning than quantifiers and *ad hoc* implicatures do. But more evidence is needed to complete the picture, and it is this that the present study contributes to.

Several papers have looked at models of strategic language use in non-cooperative contexts, with special emphasis on the question as to what happens to pragmatic inferences like quantity implicatures when the interests of interlocutors are not fully aligned or even in complete opposition (e.g., Asher & Lascarides, 2013; Franke, de Jager, & van Rooij, 2012; de Jaegher & van Rooij, 2014). An interesting puzzle which arises in this context is that if the non-cooperativeness of a conversational context is commonly known, a rational interpreter would see through any attempt of a speaker to try to mislead with an implicature: Rational conversationalists would therefore neither attempt to use, nor be fooled by, a strategically misleading implicature. Franke and van Rooij (2015) discuss how several different assumptions about cognitive limitations of language users can be combined with game-theoretic models of implicature generation in such a way that the above-mentioned puzzle dissolves. One suggestion is that language users might only have a limited capacity to reason about each other's perspective.

There is converging evidence for the view that human reasoners can engage in only a fairly limited number of theory of mind reasoning steps in strategic reasoning (e.g., Camerer, 2003; Crawford & Iriberri, 2007; Hedden & Zhang, 2002; Keysar, Lin, & Barr, 2003; Meijering, van Rijn, Taatgen, & Verbrugge, 2012; Stahl & Wilson, 1995). Models of language use that assume possibly limited theory of mind reasoning exist (e.g., Benz & van Rooij, 2007; Franke & Jäger, 2014; Pavan, 2013; Rothschild, 2013) and make predictions about adult performance in situations of strategic language use (e.g., Degen, Franke, & Jäger, 2013; Franke & Degen, 2016).

Models of iterated theory of mind reasoning also make predictions about potentially misleading uses of implicature. For example, in the system of strategic reasoning explored by Franke, de Jager, and van Rooij (2012), strategic reasoners of level-0 theory of mind sophistication do not take their co-player's perspective into account at all. Since they nonetheless pay attention to semantic meaning, the result is that level-0 speakers produce truthful utterances while level-0 listeners simply interpret all utterances as truthful without considering the speaker's possible deceptive intent. Level-(n + 1) agents, on the other hand, choose an optimal utterance or interpretation based on the belief that their co-player is a level-n agent. So, level-1 speakers assume that their listeners lack strategic capacity to reason about uncooperative language use. These speakers would therefore likely use implicatures to mislead, possibly in the hope that their listeners will not realize their intention to deceive. Conversely, level-2 listeners will anticipate an uncooperative speaker's deceptive intent but fail to consider the possibility that the speaker may be double-bluffing. However, if sophisticated speakers and listeners are uncertain about which level of sophistication their co-player is likely to have, they might use and expect multi-layered attempts of misdirection. Speakers may then try not to give any information at all, or just use random messages and listeners may also expect this behaviour. But if sophisticated speakers have sufficient evidence to classify their listener as adopting a level-0 or a level-1 strategy, they can use this expectation to inform their attempts at misdirecting them.

In the following, we therefore explore an experimental paradigm aimed at manipulating participants' beliefs about the strategic behavior of the co-player, and observing their choices of expressions to describe an object in cooperative and non-cooperative scenarios with linguistic constructions that include different types of potential implicature triggers.

2. Methods

2.1. Motivation and design

In this study, we aimed to explore how people use explicit and implicit communication in a non-cooperative context. We asked participants to play a signaling game where they had to produce verbal hints for a receiver either in a cooperative or in a competitive scenario. Our design is similar to the non-verbal paradigm of Ransom et al. (2017), but we relied on linguistic hints which include the possibility of implicatures from different kinds of triggers, and we told participants explicitly that their co-player was aware of the cooperative or non-cooperative situation. This removed the possibility of senders to select messages in such a way as to mask their non-cooperative incentives. Finally, by manipulating participants' beliefs about the strategic behavior of their co-player, we aimed to reduce the degrees of freedom in the interpretation of the behavioral choices of our participants, thereby allowing us to effectively address the central question of this work: Are language users able to strategically mislead with implicatures, triggered by different devices?

We built on previous work by Dulcinati (2018), but augment his design in order to rule out a number of potential confounding factors in the interpretation of the data, such as whether participants really understood the task, and what they believed about the interpretation of implicature triggers. Toward this end, we added feedback on the co-player's strategic behavior during an initial training phase, and collected data on each participant's beliefs about the co-player's interpretation of sentences with potential implicature triggers.

Dulcinati's (2018) study was a card-playing signaling game for two players. Each round of the game involved two cards: a winning green-bordered card and a red-bordered card. The sender knew which was the winning card and had to describe it. The receiver saw the same two cards but did not know which was the winning card. The receiver had to decide which was the correct card with the help of the description made by the sender. Participants in the experiment played the role of the sender; the receiver was a virtual player. Participants were assigned to one of three conditions: a cooperative condition or one of two competitive conditions. In the cooperative condition participants were asked to help the receiver find as many green-bordered cards as possible (a game of pure cooperation, in game theoretic terms), while in the competitive conditions their goal was to make the receiver click on as many red-bordered (losing) cards as possible (a zero sum game). The two competitive conditions differed with respect to the co-player's strategic behavior during training.

Contrary to Dulcinati (2018), we used a forced-choice design to be able to better gauge the range of answers participants could give: Participants were asked to give the receiver a hint by using one of three sentence completion options. Additionally, we aimed to have an engaging realization of the interactive task and told participants, in a form of more extreme deliberate deceit than Dulcinati (2018), that they were currently playing against another real player. By creating a stronger illusion of playing against a human player, we intended to create a more engaging task and to induce more theory of mind reasoning in the participants.[2]

Finally, we included two additional tasks in order to obtain better control of participants' likely beliefs about their interlocutor. The experiment included a training phase in which participants obtained information about the choices of the virtual co-player. By exposing different participants to different training regimes, in which the virtual co-player behaved either in a naive or a sophisticated way in the non-cooperative version of the game, we intended to induce a more concrete belief about the co-player's likely strategic behavior in the game. We also included a variant of a truth-value judgment task after the main card-playing task, with the help of which we probed participants' beliefs about their co-player's interpretation of sentences, which have an implicature reading. The data from both of these additional tasks—learning-with-feedback and truth-value judgments—were used to categorize participants into those who (a) learned their co-player's strategic behavior from training and who (b) responded semantically or pragmatically in the truth-value judgment task. We assumed that only the choices of the participants who successfully learned their co-player's strategic behavior in the training phase could give clear information regarding their attempts to mislead with an implicature in the main (non-cooperative) card-playing task. For this group of participants, we also assumed that it is more likely to see an inclination to mislead through implicature for those participants who expected their co-player to interpret implicature-triggering sentences in a more pragmatic way, rather than just literally.[3]

2.2. Participants

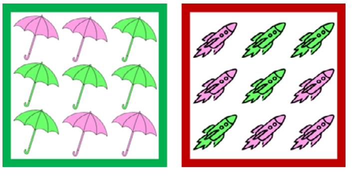
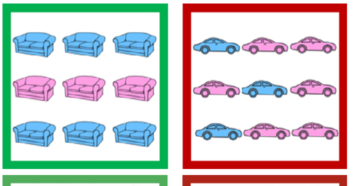
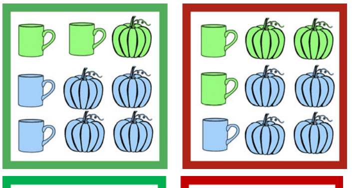
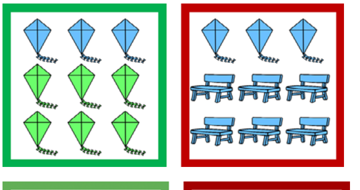
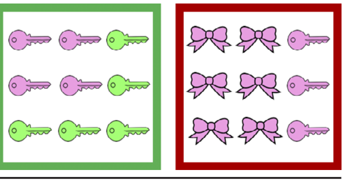
We collected data from 103 participants (70 female, 33 male; mean age 36.6 years, $SD = 12.6$ years) via the crowd-sourcing platform Prolific. We recruited participants who satisfied the following pre-conditions, using Prolific's built-in screening tools: English (any variety) as a first language, minimum of 50% acceptance rate on a minimum of 10 previous experiments on Prolific, and indicated willingness to engage in studies which apply deliberate deceit. Each participant was paid £2.50 for compensation. The experiment description also promised a potential bonus payment of up to £0.75 based on the degree of success in the card-playing game. During the debriefing at the end of the experiment, participants were informed that they had not played against a human co-player. They were also informed that, contrary to what had been advertised in the instructions, they would receive the full bonus payment irrespective of their answers.

2.3. Materials (card-playing task)

The material for a single trial of the card-playing task consisted of a picture depicting two cards—one green-bordered and the other red-bordered—as well as an initial sentence string and three sentence continuations, out of which the participants had to choose one.

The initial sentence string was always the same: “The green card is the card where . . ..” The three options for completing this sentence differed for the five sentence-type conditions, which were *all*, *none*, *number*, *ad hoc* and *some* (see Table 1 for an example of each condition type). The three sentence completion options always corresponded to the categories “green,” “red,” and “opt out.” The “green” sentence continuation would result in a sentence being literally true solely of the green-bordered card in the *all* and *none* conditions. In the other (implicature) conditions—*number*, *ad hoc,* and *some*—the “green” sentence completion was literally true of both cards, but referred uniquely to the green-bordered card under the sentence’s implicature reading. The same applied, mutatis mutandis, for the “red” option. The “opt out” option resulted in a sentence which was either true of both cards or true of neither card. There were nine different trials for each

Table 1
Examples of material for the different conditions in the card-description task. The green-bordered card is on the left in each row of the “Picture” column.

Type	Option	Completion	Picture
All	Green Red Opt out	All of the objects are umbrellas All of the objects are rockets Some of the objects are clowns	
None	Green Red Opt out	All of the objects are cars All of the objects are sofas Some of the objects are shivering	
Number	Green Red Opt out	Five of the objects are pumpkins Five of the objects are blue Seven of the objects are dragons	
Ad hoc	Green Red Opt out	All of the objects in the top row are blue All of the objects in the top row are kites Some of the objects are friendly	
Some	Green Red Opt out	Some of the objects are pink Some of the objects are keys Some of the objects are frogs	

of the *all* and *none* conditions, and six for each of the other three (implicature) conditions—*number*, *ad hoc*, and *some*.

2.4. Materials (truth-value judgment task)

The truth-value judgment task presented a sentence on each trial with a set of two-alternative forced-choice answers: "true" and "false." Sentence material was inspired by, and partly borrowed from, Noveck (2001). A total of 60 sentences instantiated three conditions: Bizarre, False, and Implicature. Examples and numbers of instances are given in Table 2. Sentences in the Bizarre condition were expected to yield "false" judgments, as they involved a nonsensical element such as a type mismatch. Sentences in the "true" condition were expected to receive "true" judgments based on common world knowledge. Sentences in the Implicature condition would be judged true on a literal reading, but false under an implicature reading.

2.5. Procedure

The experiment lasted about 20 min on average and consisted of two main parts: the card-playing task and the truth-value judgment task. Each participant was randomly assigned to one of three conditions at the beginning of the experiment: the Cooperative, the Unstrategic, or the Strategic condition. These conditions influenced the way in which the card-playing task was set up, as well as the type of feedback the participant would receive during the initial training.

The experiment began by providing general instructions. Participants were informed that there were two parts to the experiment and that both were equally important. It was stressed that this experiment involved playing with another human player, and that therefore participants should pay particular attention to the instructions. Since some of the tasks required normal color vision to see the—"green" and "red"—card borders,

Table 2
Examples of materials for the truth-value judgment task

Condition	Type	Instances	Example
Bizarre	All	5	All birds are telephones
Bizarre	Most	5	Most elephants have glasses
Bizarre	Number	5	The moon has two ears
Bizarre	Some	5	Some stores are made of bubbles
Implicature	Most	5	Most horses are mammals
Implicature	Number	5	A human hand has two fingers
Implicature	Some	5	Some trouts are fish
True	All	10	All hammers have a handle
True	Most	5	Most people know their parents
True	Number	5	A week has seven days
True	Some	5	Some drinks are made of chocolate

participants saw three plates from the Ishihara color-blindness test (Ishihara, 1917) following the general instructions.

Participants were then made familiar with the card-playing game which constituted part 1 of the experiment. Irrespective of the condition each participant was assigned to, the card-playing task was introduced as an interactive game with another real player. Participants were told that their role was that of the *describer* and that the other player was to play as the *guesser*. Participants learned that there were always two cards in each trial, one bordered by green, the other by red, referred to as the "green card" and "red card," respectively. They were told that as a *describer* they could see these colors, but not the guesser. The guesser would see the same cards without colored borders and possibly in a different spatial arrangement. The guesser's task was to choose the card with the green border, based on the describer's description of it. As describers, participants were to select one out of three continuations to complete sentences starting with "The green card is the card where . . .." To ensure a good understanding of the instructions, participants saw an example screenshot of the task from both the guesser's and the describer's points of view (see Fig. 1 for examples of the describer's and guesser's task, as presented during the instructions). Participants assigned to the Cooperative condition were instructed to provide a continuation which would enable the guesser to choose the green card. Participants assigned to the other two (non-cooperative) conditions were to make the guesser choose the red card. We told participants that they would receive a bonus payment proportional to the number of rounds they played successfully, that is, where the guesser managed to pick out the intended card: green in the Cooperative condition and red in non-cooperative conditions.

Before the main test phase of part 1, there were two practice rounds of four trials each. In the first of these, the participants practiced four trials in the role of the guesser, without any feedback. They were then told that they were being partnered with another player. Although no actual matching took place, the experiment simulated this process by making participants wait for 9 seconds before a message "Pairing successful!" appeared and participants could proceed.

During the second training round, participants played 18 trials in the describer role, exactly as they would in the subsequent test phase. These training sequences used the *all* and *none* items from the main test phase, presented in randomized order. Unlike what would happen in the test phase, participants received feedback for each round after a brief delay, programmed to simulate the co-player making a choice. The co-player's simulated behavior was dependent on the condition (Cooperative, Strategic, and Unstrategic), except when participants chose the "opt out" option, in which case the co-player was simulated to select a random card. In the Cooperative and Unstrategic conditions, the simulated co-player always interpreted the sentence literally, that is, choosing the green card when a "green" continuation had been selected and the red card when a "red" continuation had been chosen by the participant. By contrast, in the Strategic (non-cooperative) condition, the simulated co-player always chose the reverse, for example, the red card after a selection of "green." Therefore, the Strategic condition meant to induce the belief in participants that their *guesser* co-player played like a rational agent who consistently

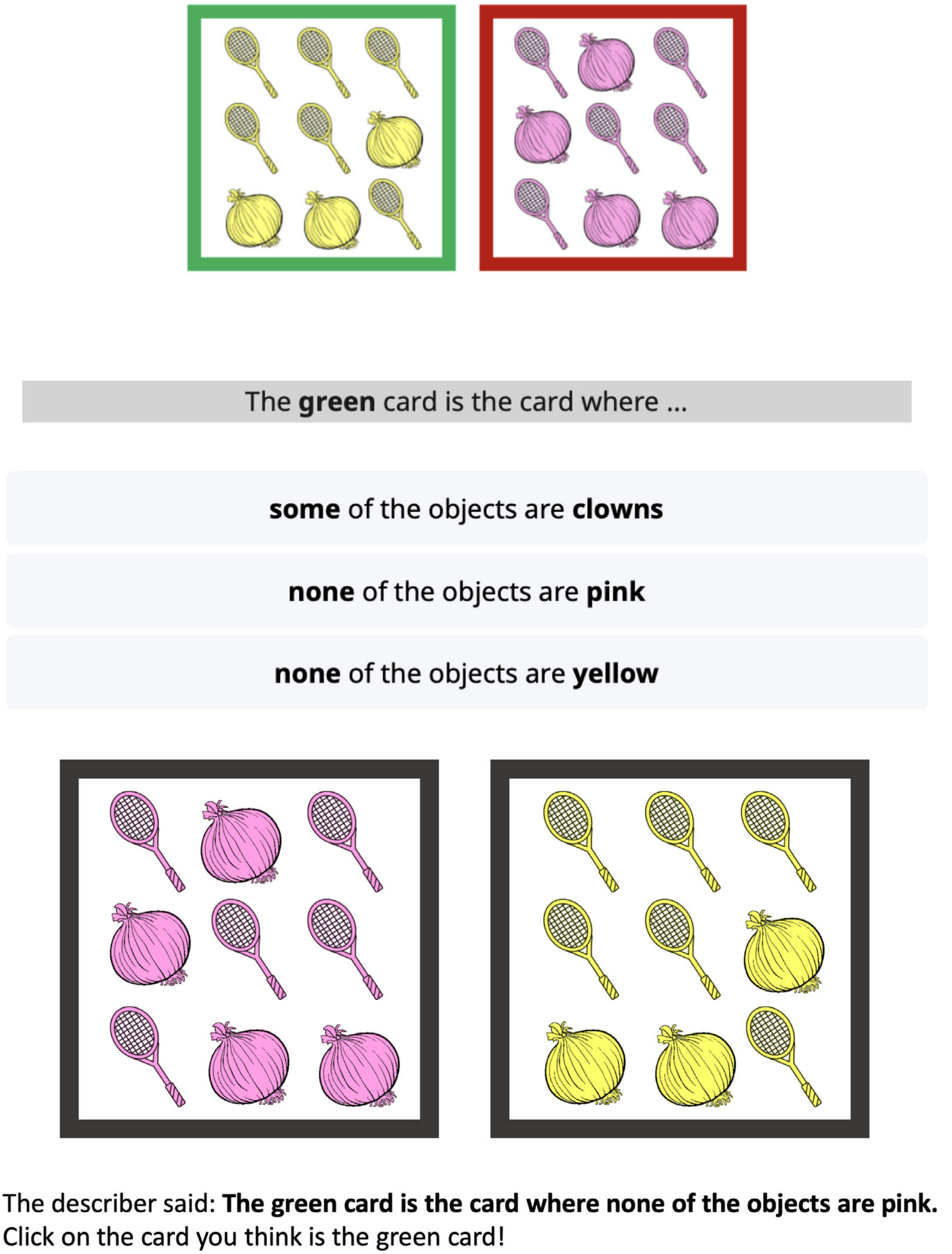


Fig. 1. Example trials for the main card description task, as shown to participants during the instructions. Top: Example of the describer's task. Bottom: Example of the guesser's task.

chose the best response to each description, based on the assumption of a non-strategic describer behavior (corresponding to a level-1 theory of mind strategy). The Unstrategic co-player behavior instead tried to induce the belief in participants that their *guesser* co-player might be oblivious to the non-cooperative nature of the game (corresponding to a level-0 theory of mind strategy, i.e., without factoring in the describer's possibly divergent interests in the game).

Following the training phase, the main test phase of part 1 presented all of the 36 trials in random order. Participants received no feedback on the choices of the virtual co-player, but experienced short pauses of random duration (between 2 and 4 s) before getting a response to maintain the illusion that they were playing against a real co-player.

The card game was followed by a self-assessment in which participants adjusted a slider ranging from 0% to 100% to indicate the percentage they believed to have played successfully during the main test phase.

The second part of the experiment began with general instructions. Participants were informed that we wanted to know what they thought about their co-player's interpretation of certain sentences. The explicit emphasis on participants' beliefs about the co-player's interpretation rather than their own was deliberate, since, strictly speaking, the beliefs about co-player interpretation are those that should matter for strategic choice. Admittedly, it may be natural to assume at least *prima facie* that one's interlocutor interprets sentences exactly in the same way as we would ourselves. Even if participants did not believe that their co-player's interpretation might diverge from their own, their answers to these questions should nonetheless indicate the strategic reasons for their choices in the card-description task.

Participants read 60 sentences presented in random order. They indicated whether they thought their "partner" would consider each sentence true or false by clicking on the "true" or "false" button. A short 4-second interval was built in before the answer buttons appeared, to allow the participants time to read and process the sentence. Initial piloting of the sentence material revealed that without this initial waiting time, the number of pragmatic responses was close to zero, that is, "false" judgments of sentences of the *some*, *most,* and *number* type in the Implicature. This suggests that, without the added pause, participants might have rushed through the material too quickly to process potential implicatures.

The experiment ended with a debriefing, in which participants were informed about the deliberate deceit and told that they would receive the full bonus payment of £0.75 irrespective of their answers on the trials. Finally, participants could, if they wanted, once more provide information on their native language, gender, and education level and leave any comments that they might have about the experiment.

3. Results

The randomized allocation of participants to conditions resulted in 39 participants playing the Cooperative condition, 33 for the Unstrategic and 31 for the Strategic condition.

Despite occasional minor mistakes (e.g., mistaking a "7" for a "1"), no participant blatantly failed the color blindness test, and so nobody was excluded based on this.

The co-player behavior exhibited during the training phase with feedback would allow participants, if they were able to learn the behavior of the co-player, to play a unique optimal winning move. The option "green" was the optimal choice in the Cooperative and Strategic condition. The option "red" was optimal for the Unstrategic condition. We looked at each participant's proportion of choices of the optimal option (out of 18 trials). The mean proportion was 0.99 in the Cooperative condition, 0.64 in the Strategic condition, and 0.52 in the Unstrategic condition. We considered any participant with a proportion of optimal choices of at least 0.5 as having learned successfully during training. This resulted in all 39 participants of the Cooperative condition being categorized as successful learners, as well as 19 out 33 participants in the Unstrategic and 19 out of 31 in the Strategic condition.

All items in the True and Bizarre conditions of the truth-value judgement could clearly be identified as true or false based on common world knowledge. We excluded seven participants from further analysis who did not surpass a threshold of 70% correct answers on these items. We use the answers to the *some* and *most* condition to probe into participants' likely beliefs about their co-player's disposition to compute scalar implicatures. We classified a participant as a pragmatic type if the proportion of "false" judgements in the *some* and *most* condition exceeds 0.5; otherwise the participant is classified as a semantic type. A total of 36 participants was classified as a pragmatic type in this way.

Fig. 2 shows the proportions of choice options, split by semantic vs. pragmatic type and successful vs. unsuccessful learning. It is apparent that the aggregate behavior of participants with successful learning (subplots A and B) differs strikingly from those who did not learn to play well against their co-player during training with feedback (subplots C and D). The modal choice for successful learners is the winning move for each respective condition. In contrast, there is no clear modal choice or otherwise visually salient pattern in the aggregate data of unsuccessful learners.

As planned, we focus our attention and analyses on the data from participants with successful training results. Our main research questions concern the circumstances which influence the probability that participants choose the winning move. We therefore ran a logistic regression analysis, predicting the log-odds of winning-move choices against non-winning-move choices with predictors CO-PLAYER-TYPE (levels: "cooperative," "unstrategic," and "strategic"), CONSTRUCTION (levels: "all," "none," "number," "ad hoc," and "some"), and IMPLICATURE-TYPE (levels: "semantic," "pragmatic"). We used the *brms* package (Bürkner, 2017) and statistical programming language R (R Core Team, 2017), to fit a hierarchical logistic regression with by-item random intercepts using Bayesian inference of a posteriori credible values of model parameters. We fixed wide, largely uninformative priors for fixed-effects coefficients, using a normal distribution with mean 0 and standard deviation 30. We collected 2,000 samples from each of 4 chains after a warm-up of 1,000 samples and ensured convergence by inspection of traceplots and R-hat statistics. The regression model used the default dummy coding of factors (with "cooperative," "all" and "pragmatic" as reference levels), but Bayesian inference allows us to

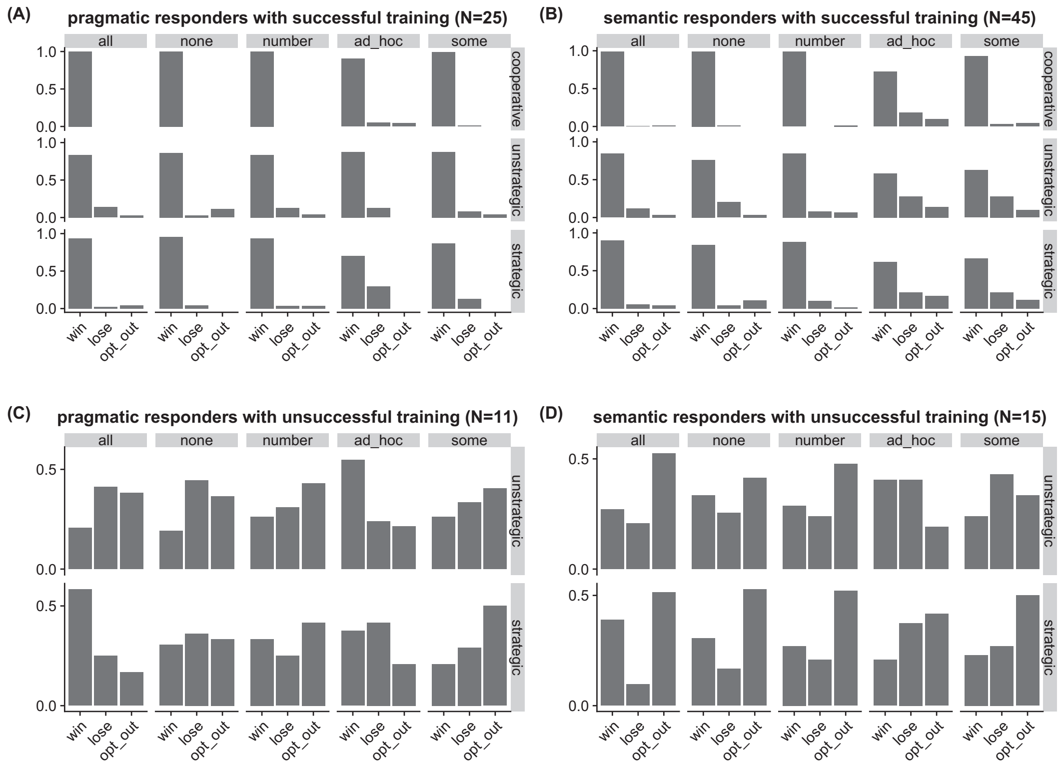


Fig. 2. Proportion of choice options selected by participants who learned successfully during the training-with-feedback session (A and B) or did not (C and D), and who were classified as pragmatic type (A and C) or as semantic type (B and D) based on their behavior in the truth-value judgement task.

compute approximate posteriors for any relevant comparison of (sets of) design cells from the obtained samples. In the following, we will say that the data, when analyzed through this model, provide evidence in favor of the assumption that a combination of factor levels i has a credibly higher proportion of winning-move choices than a combination of factor levels j, if the posterior level of credence $P(\delta_{ij} > 0|D)$, given data D, for the difference δ_{ij} between log-odds inferred for the estimated coefficients for i minus that for j is estimated to be at least 0.95. We also report below the 95% credible interval of the posterior distribution of $P(\delta_{ij}|D)$ (Kruschke, 2015).

We start by investigating the main effects that different factor levels have on the average log-odds of winning-move choices. As for main effects of IMPLICATURE-TYPE, the probability of winning-move choices is credibly higher for pragmatic types than semantics types ($P(\delta_{ij} > 0|D) \cong 1$, CI = [0.6717; 2.685]). Regarding main effects of CO-PLAYER-TYPE, we find that the cooperative condition had a higher probability of winning-move choices than the strategic co-player condition ($P(\delta_{ij} > 0|D) \cong 1$, CI = [2.637; 8.635]), and that the latter had higher winning-move choice probability than the unstrategic co-player condition ($P(\delta_{ij} > 0|D) \cong 0.9722$, CI = [−0.0302; 0.8354]). For

CONSTRUCTION, we find no credible difference between any of "all," "none," and "number," but find that "some" has a lower probability of winning-move choices than "none," which has the lowest observed mean of winning-move choices among the former three ($P(\delta_{ij} > 0|D) \cong 0.9995$, CI = [0.552; 6.194]). Finally, "some" has a credibly higher rate of winning-move choices than "none" ($P(\delta_{ij} > 0|D) \cong 0.9992$, CI = [0.3719; 1.667]).

We look next at the effect of CONSTRUCTION on the winning-choice rates for pragmatic responders. For the Cooperative condition, we find that "all," "none," and "number" are not credibly different, but that "some" is credibly lower than all of the former ($P(\delta_{ij} > 0|D) \cong 0.9704$, CI = [−1.541; 30.92], when compared against "number") and that "ad hoc" is in turn credibly lower than "some" ($P(\delta_{ij} > 0|D) \cong 0.9975$, CI = [0.4731; 5.582]). For the Unstrategic co-player condition, we find no credible difference between any construction type at all. For the Strategic co-player condition, only "ad hoc" is credibly smaller than any one of "all," "number," and "none" ($P(\delta_{ij} > 0|D) \cong 0.994$, CI = [0.3007; 3.838], when compared against "number").

The effects of CONSTRUCTION for semantic responders are as follows. For the Cooperative condition, all of "all," "none," and "number" are not credibly different, but "some" has credibly lower winning-choice rates than all of the former ($P(\delta_{ij} > 0|D) \cong 0.9945$, CI = [0.23; 3.662], when compared against "number"), and "ad hoc" has in turn credibly lower winning-choice rates than "some" ($P(\delta_{ij} > 0|D) \cong 1$, CI = [0.9035; 2.43]). As for the Unstrategic co-player condition, we find no credible difference between "all," "none," and "number," but find that "some" has credibly lower winning-choice rates than all of the three former ($P(\delta_{ij} > 0|D) \cong 0.9991$, CI = [0.3975; 2.056], when compared against "number"). Unlike in the Cooperative condition, the difference between "some" and "ad hoc" is not credibly different in the Unstrategic co-player condition ($P(\delta_{ij} > 0|D) \cong 0.6984$, CI = [−0.536; 0.8637]). Finally, in the Strategic co-player condition, we find the exact same pattern as for the Unstrategic condition. Again, "all," "none," and "number" are not credibly different, but "some" has credibly lower winning-choice rates ($P(\delta_{ij} > 0|D) \cong 0.9975$, CI = [0.4343; 2.404], when compared against "number"), while "ad hoc" and "some" are not credibly different.

In sum, we find that "number" consistently patterns with "all" and "none," while "some" and "ad hoc" are almost always associated with credibly lower winning-choice rates. The only exceptions to this pattern are found for participants classified as pragmatic responders in the non-cooperative conditions: In the Unstrategic condition, we find no difference between any constructions, while in the Strategic condition participants chose the winning-move for "some" to an extent that is not credibly different from that of "all," "none," and "number."

4. Discussion

Our participants played in the sender role of a signaling game in which they were either helping or competing against a virtual co-player in the receiver role. Their task was to complete descriptions that could help their co-player choose the winning card from a pair of cards. Some items pushed signalers to convey this hint via assertion and others via implicature (numerals, *some*, and *ad hoc*). Before playing the game, participants played a training phase where they could learn the strategy their co-player was using to interpret their instructions. After the game, participants completed a truth-value judgment task in which they indicated how they thought their co-player in the game would have interpreted a series of statements that can give rise to implicatures.

Unless we explicitly indicate otherwise, we will focus our attention on participants who completed the training phase of the card game successfully. This criterion can be used as a strong indicator that they were paying attention to the instructions, understood them, and followed them.[4] Indeed, these participants' responses to the main card-game task form a fairly coherent pattern of behavior. By contrast, the picture emerging from the responses of participants who were unsuccessful during learning is erratic, as one would expect if these participants were unwilling or unable to follow the instructions, or hold a specific hypothesis about the co-player's behavior.

The participants were further divided into semantic and pragmatic responders depending on how they responded in the truth-value judgment task. From the participants who performed successfully during training, 25 participants were classified as pragmatic responders, 45 as semantic responders. The former class, pragmatic responders with successful training, is the theoretically most interesting group for our purposes, because it is here, if at all, that we would expect to see misleading uses of implicatures. Nonetheless, a visual inspection of the data in Fig. 2 (A and B) suggests a lot of similarities in how pragmatic and semantic responders performed on the card-game task. The distinction between pragmatic and semantic responders might in fact yield interesting additional insights; we will thus consider all successful training participants in our discussion of the findings.

Both semantic and pragmatic participants overwhelmingly chose the winning description in the Cooperative condition. This was expected for pragmatic responders, given that their aim was to help their co-player identify the bonus card. It may be more surprising for semantic responders who expected their co-players to infer implicatures in the context of the game, despite indicating in the truth-value judgment task they did not expect them to infer implicatures "in general." The truth-value judgement task and the card game are very different paradigms and it seems plausible that the truth-value judgment task leads to underestimate some participants' abilities to take into account implicatures. Indeed, the rate of implicature understanding is generally fairly low when it is assessed by a truth-value judgment task (e.g., 59% in Noveck, 2001; ≤60% in Bott & Noveck, 2004; 55% in Mazzarella, Reinecke, Noveck, & Mercier, 2018). Ultimately, the semantic responders performance in the Cooperative condition suggests that we should probably be cautious with the distinction we drew between semantic and pragmatic responders based on the truth-value judgment task. Importantly, all participants who passed the training phase demonstrated a rational behavior in the Cooperative condition. While Dulcinati (2018) also found that participants used the three types of implicatures in the Cooperative condition of his production study, their pattern of responses was much less clear. This may have been due to participants who did not follow the instructions or had no clear hypothesis about the co-player's behavior, or due to the more open-ended nature of his production study. Our paradigm is more constrained and we added a training phase, which allows us to focus on participants who definitely understood the strategic aspects of the game. Their behavior in the Cooperative condition therefore provides a good baseline to assess their strategies in the two competitive conditions.

Pragmatic as well as semantic responders took into account the competitive nature of the card game in the Strategic and Unstrategic conditions and exhibited different behaviors than in the Cooperative condition. Fig. 2 (A and B) suggests that all types of responders with successful training used implicatures to deceive their co-players in the Strategic and Unstrategic conditions. This behavior is enhanced in pragmatic participants, who find implicatures more accessible and presumably find it easier to recruit them to deceive their co-player. Again, this general finding echoes Dulcinati's (2018) production study, in which participants used the three types of implicatures to deceive their co-player in the competitive version of his signaling card game, but the present results are far more crisp and less likely affected by confounding factors.

Our production findings seem in line with the two studies conducted on implicature comprehension in non-cooperative contexts, which employed a game setting as we do in this study (Dulcinati, 2018; Pryslopska, 2013). In both studies, participants interpreted descriptions given by an uncooperative speaker and had to decide which "card" or situation the speaker was describing by choosing an option from a set of alternatives including a pragmatic interpretation and a semantic interpretation of the description. Both studies found that participants were less likely to choose the option that matches the pragmatic interpretation of the description when the speaker was uncooperative, compared to when the speaker was cooperative. One key difference with this study is that neither of these two comprehension studies allowed participants to clearly infer the speaker's strategy in the game. This means that their participants may have expected a more sophisticated strategy whereby the speaker's implicatures are sometimes true and sometimes false. Indeed, in Dulcinati's (2018, comprehension) study, participants interpreted both assertions and implicatures as false half of the time and true half of the time. The findings from these two comprehension studies are consistent with our production results. Their "hearers'" behavior mirrors that of our "speakers" and they illustrate how listeners expect an uncooperative speaker's behavior to deviate from the cooperative speaker's consistently truthful use of implicatures.

Interestingly, in our study, participants chose the winning description more often in the Strategic than in the Unstrategic condition. As we had anticipated, participants performed the best in the Cooperative condition. This is the simplest condition since it involves no deceit, no complex perspective-taking or reasoning over their co-player's strategy: Participants described what they knew was the winning card. The difference we observe between the Strategic and the Unstrategic conditions, on the other hand, might seem counterintuitive, since the Unstrategic condition involves simple deceit, while the Strategic condition involves double-bluffing—the co-player sees through the deceit and chooses the opposite card from the one misleadingly described ("green" when "red" is hinted to). Yet, the participants' behavior makes sense when one considers the actions they have to perform on repeated trials. The theory of mind reasoning involved in the Strategic condition for the participant to select the winning description is the most complex. They have to (a) pay attention to the “green card”, (b) intend for their co-player to pick up the “red card”, but (c) considering their co-player’s known strategy (from the training phase), and (d) describe the “green card” after all. While the reasoning is complex, this eventually requires a relatively simple action: describe the “green card”, precisely what they would have done in the Cooperative condition. Once it has been thought through, implementing this strategy over a series of trials is straightforward. The Unstrategic condition, on the other hand, involved a simpler misleading strategy: describe the “red” losing card since the co-player is unstrategic and picks up the card communicated to them (as indicated by the training phase). While this reasoning involves fewer levels of meta-representation than that of the Strategic condition, its implementation is more difficult. Describing the “red” losing card requires ignoring the pull toward the “green” winning card, thereby putting participants inhibitory control to the test in every trial.[5]

Participants did not use the three types of implicature triggers in an identical fashion. All were used to produce implicatures, but not in equal measure. The most striking effect is the difference between numerals and the two other types of implicature triggers. A main effect indicates that the choices of winning descriptions for numerals patterns with those of *all* and *none* control items overall—there is also no difference between them in the Cooperative condition. Participants, thus, used the exact interpretation of numerals as reliably as truth conditional content. This suggests that, if the *exact* interpretation of numerals is an inference, in our study it was as available as the semantic meaning of the quantifiers *all* and *none*. Indeed, while participants used numeral items in a comparable way to the control items, they did so differently from either other implicature triggers, *some,* or *ad hoc*. Dulcinati (2018) also found in his production study that participants treated numerals’ exact reading in the same way as the meaning of literal controls, in contrast to *ad hoc* and scalar (*most*) implicatures. This pattern of results echoes that of Weissman and Terkourafi (2019), who found that false numeral implicatures were considered as lies, unlike other implicature types including those linked to scalar quantifiers. All these findings suggest that the *exact* interpretation of numerals is part of their truth-conditional meaning (Breheny, 2008; Carston, 1998; Geurts, 2006; Kennedy, 2015) and not an implicature (Gadzar, 1979; Horn, 1972; Levinson, 2000)—a conclusion supported by recent experimental evidence on the processing of *some* and numerals (Marty, Chemla, & Spector, 2013).

Numerals are set aside by the specific behavior—akin to that of the literal meaning—they give rise to. Yet, while the responses elicited by *some* and *ad hoc* implicatures appear similar and are clearly distinct from those of *all*, *none,* and numerals, they are not identical to each other. Our regression analysis suggests a credible difference in the Cooperative condition for both pragmatic and semantic responders, where the scalar trigger *some* prompted more winning description choices than did the *ad hoc* implicature. The same difference was found for pragmatic responders in the Strategic condition, but not for pragmatic responders in the Unstrategic condition, or semantic responders in either of the two non-cooperative settings, where *some* and *ad hoc* were not credibly different. Participants’ reaction to *some* was therefore sometimes more similar to literal (and numeral) items than *ad hoc* triggers were. This line of reasoning suggests that scalar implicatures triggered by *some* are no more difficult, if not perhaps easier to access than those triggered by *ad hoc scales*. This can be interpreted in two different ways. It can be viewed as an argument in favor of a divide between two types of implicatures, with *ad hoc* being particularized implicatures and *some* quantity implicatures, a type of pragmatic inference closer to the truth-conditional content of the utterance (e.g., a Generalized Quantity Implicature for Horn, 1972 or Levinson, 2000; a "primary pragmatic process" for Recanati, 2003).[6] We feel the difference we find between *some* and *ad hoc* implicature triggers might be insufficient to justify such a strong theoretical distinction—although it is certainly compatible with it. An additional reason for caution is that Dulcinati's (2018) findings diverge from ours here on this account: He found that scalar implicature with *most* elicited the same responses as *ad hoc* implicature triggers. We therefore favor a more general interpretation of our results: Quantifier and *ad hoc* triggers exhibit an accessibility variability found elsewhere between implicature triggers (Doran et al., 2012; Weissman & Terkourafi, 2019—note that this variability does not correspond neatly to a divide between *particularized* and *generalized* implicatures or *primary* and *secondary* processes).

In conclusion, we found that cooperative, but also uncooperative, speakers can produce implicatures at least in the kind of signaling card game we set up in our study. Furthermore, speakers will adapt their use of (false) implicature to the listeners' typical response behavior. They exploit implicatures to deceive, but not as consistently as literal meaning (lies) and not equally across the board of implicature triggers. These findings fit well with those of other studies: in production (Dulcinati, 2018), comprehension (Pryslopska, 2013), or using other paradigms (Weissman & Terkourafi, 2019) where (false) implicatures are used and understood in uncooperative contexts, or treated as lies by listeners, but not to the same extent as literal meaning—a distinction also found in the assessment of speaker commitment (Mazzarella et al., 2018). Additionally, what seems to emerge from our results is a scale across numeral, *some,* and *ad hoc* items, where the exact reading of numerals is closest to truth conditional content, *ad hoc* scalar implicature furthest, and scalar implicatures of *some* somewhere in the middle. We argue that the qualitative distinction on this scale is that between numerals, on the one hand, and *some* and *ad hoc* scales, on the other, while the difference between *some* and *ad hoc* is due to different availability of the upper bound reading between these two implicatures. Further experimental research on the use of implicatures in non-cooperative situations is needed to better understand which factors influence the speakers' decision to communicate something implicitly or explicitly, as well as the nature of the implicatures themselves and the role of cooperation in their derivation.

Notes

1. We assume here an explicit/implicit divide corresponding to the distinction between Grice's *what is said* and what is *implicated*. However, see Carston (2002, 2009) and Recanati (2004) for different perspectives.

2. Another difference between our design and that of Dulcinati (2018) was the substitution of the scalar implicature trigger *most* for the trigger *some*. *Most* might allow for unintended additional readings, while *some* is less ambiguous and better researched regarding its scalar implicature behavior in cooperative contexts.
3. All experimental material, the experiment code itself, the data and analysis scripts are available at the OSF repository with https://doi.org/10.17605/OSF.IO/497F3.
4. We should emphasize that focusing on the subset of participants with successful training is a harsh measure aiming to obtain a data set with as few as possible of the potentially confounding factors that tainted the interpretation of Dulcinati's (2018) results. As pointed out by an anonymous reviewer, participants who were not classified as successful learners here might, after all, have understood the task as well, and might have behaved very strategically, namely using the training phase not only to learn about the co-player's behavior but also trying to disguise their own signaling strategy by deliberate randomization. It is hard, if at all possible, to infer a participant's strategy from their behavioral choices alone. This brings us back to the main motivation behind the present design and analysis, which is to address, as unambiguously as possible, whether participants can use implicature triggers strategically in uncooperative contexts to mislead by false implicature.
5. Note that in the Cooperative and Strategic conditions, the winning description is that of the "green card", while in the Unstrategic condition, it is the description of the "red card". It is therefore possible to imagine that in the Strategic condition, participants simply revert to the easier Cooperative condition strategy. But since the Unstrategic participants considered here are those who successfully completed the Unstrategic condition training, we can confidently rule out this simpler explanation.
6. The more extreme version of this line of thought maintaining that implicatures arising from lexicalized scales are default meanings computed automatically (e.g., Chierchia, 2004; Levinson, 2000) is not supported by the data. If this were the case, participants should have treated *some* in the same way as literal triggers (and numerals).

References

Asher, N., & Lascarides, A. (2013). Strategic conversation. *Semantics and Pragmatics*, *6*, 1–22.

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *188*, 87–96.

Benz, A., & van Rooij, R. (2007). Optimal assertions and what they implicate. *Topoi*, *26*, 63–78.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*(3), 437–457.

Breheny, R. (2008). A new look at the semantics and pragmatics of numerically quantified noun phrases. *Journal of Semantics*, *25*(2), 93–139.

Bull, P. (2016). Theresa May has a very special technique for avoiding questions. *The Conversation*. Retrieved from theconversation.com.

Bürkner, P. (2017). *brms*: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.

Carston, R. (1998). Informativeness, relevance and scalar implicature. In R. Carston & S. Uchida (Eds.), *Relevance theory: Applications and implications* (pp. 179–236). Amsterdam: John Benjamins.

Carston, R. (2002). *Thoughts and utterances: The pragmatics of explicit communication*. Oxford, UK: Blackwell.

Carston, R. (2009). The explicit/implicit distinction in pragmatics and the limits of explicit communication. *International Review of Pragmatics*, *1*(1), 35–62.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and beyond* (pp. 39–103). New York: Oxford University Press.

Chierchia, G., Fox, D., & Spector, B. (2012). Scalar implicature as a grammatical phenomenon. In: C. Maienborn, K. vonHeusinger, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (vol. 3, pp. 2297–2331). Berlin: Mouton de Gruyter.

Coleman, L., & Kay, P. (1981). Prototype semantics: The English word lie. *Language*, *57*(1), 26–44.

Crawford, V. P., & Iriberri, N. (2007). Fatal attraction: Salience, naïveté, and sophistication in experimental "hide-and-seek" games. *The American Economic Review*, *97*(5), 1731–1750.

de Jaegher, K., & van Rooij, R. (2014). Game-theoretic pragmatics under conflicting and common interests. *Erkenntnis*, *79*, 769–820.

Degen, J., Franke, M., & Jäger, G. (2013). Cost-based pragmatic inference about referential expressions. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 376–381). Berlin: Cognitive Science Society.

Doran, R., Baker, R. E., McNabb, Y., Larson, M., & Ward, G. (2009). On the non-unified nature of scalar implicature: An empirical investigation. *International Review of Pragmatics*, *1*(2), 211–248.

Doran, R., Ward, G., Larson, M., McNabb, Y., & Baker, R. E. (2012). A novel experimental paradigm for distinguishing between what is said and what is implicated. *Language*, *88*(1), 124–154.

Dulcinati, G. (2018). Cooperation and pragmatic inferences. Doctoral dissertation, University College London, London, UK.

Dulcinati, G., & Pouscoulous, N. (2017). Quantity implicatures in a competitive game. Talk presented at XPRAG, Cologne, Germany.

Dynel, M. (2011). A web of deceit: A neo-Gricean view on types of verbal deception. *International Review of Pragmatics*, *3*(2), 139–167.

Dynel, M. (2015). Intention to deceive, bald-faced lies, and deceptive implicature: Insights into Lying at the semantics-pragmatics interface. *Intercultural Pragmatics*, *12*(3), 309–332.

Franke, M., de Jager, T., & van Rooij, R. (2012). Relevance in cooperation and conflict. *Journal of Logic and Computation*, *22*(1), 23–54.

Franke, M., & Degen, J. (2016). Reasoning in reference games: Individual- vs. population-level probabilistic modeling. *PLoS ONE*, *11*(5), e0154854.

Franke, M., & Jäger, G. (2014). Pragmatic back-and-forth reasoning. In S. Pistoia Reda (Ed.), *Semantics, pragmatics and the case of scalar implicatures*, Palgrave Studies in Pragmatics Language and Cognition, chapter 7 (pp. 170–200). New York: Palgrave MacMillan.

Franke, M., & van Rooij, R. (2015). Strategies of persuasion, manipulation and propaganda: Psychological and social aspects. In J. van Benthem, S. Gosh, & R. Ver-brugge (Eds.), *Models of strategic reasoning: Logics, games and communities* (pp. 255–291). Heidelberg: Springer.

Gadzar, G. (1979). *Pragmatics: Implicature, presupposition and logical form*. New York: Academic Press.

Geurts, B. (2006). Take "five": the meaning and use of a number word. In S. Vogeleer & L. Tasmowski (Eds.), *Non-definiteness and plurality* (pp. 311–329). Amsterdam: Benjamins.

Geurts, B. (2010). *Quantity implicatures*. Cambridge: Cambridge University Press.

Grice, H. P. (1989). *Studies in the way of words*. Cambridge: MA: Harvard University Press.

Hardin, K. J. (2010). The Spanish notion of lie: Revisiting Coleman and Kay. *Journal of Pragmatics*, *42*(12), 3199–3213.

Hedden, T., & Zhang, J. (2002). What do you think I think you think?: Strategic reasoning in matrix games. *Cognition*, *85*(1), 1–36.

Hirschberg, J. (1991). *A theory of scalar implicature*. New York: Garland Press.

Horn, L. (1972). On the semantic properties of logical operators in English. UCLA dissertation, distributed by Indiana University Linguistics Club, 1976.

Horn, L. R. (1989). *A natural history of negation*. Chicago, IL: University of Chicago Press.

Huang, Y., & Snedeker, J. (2009). On-line interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology*, *58*, 376–415.

Huang, Y. T., Spelke, E., & Snedeker, J. (2013). What exactly do numbers mean? *Language Learning and Development*, *9*(2), 105–129.

Isenberg, A. (1973). *Aesthetics and theory of criticism: Selected essays*. Chicago, IL: University of Chicago Press.

Ishihara, S. (1917). *Tests for color-blindness*. Handaya, Tokyo: Hongo Harukicho.

Katsos, N. (2009). Evaluating under-informative utterances with context-dependent and contextindependent scales: Experimental and theoretical implications. In U. Sauerland & K. Yatsushiro (Eds.), *Experimental semantics and pragmatics* (pp. 51–73). Basingstoke, UK: Palgrave.

Katsos, N., & Cummins, C. (2010). Pragmatics: from theory to experiment and back again. *Language and Linguistics Compass*, *4*(5), 282–295.

Kennedy, C. (2015). A “de-Fregean” semantics (and neo-Gricean pragmatics) for modified and unmodified numerals. *Semantics & Pragmatics*, *8*(10), 1–44.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25–41.

Kruschke, John E. (2015). *Doing Bayesian data analysis*. Burlington, MA: Academic Press.

Levinson, S. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Mahon, J. E. (2007). A definition of deceiving. *International Journal of Applied Philosophy*, *21*(2), 181–194.

Marty, P., Chemla, E., & Spector, B. (2013). Interpreting numerals and scalar items under memory load. *Lingua*, *133*, 152–163.

Mazzarella, D., Reinecke, R., Noveck, I., & Mercier, H. (2018). Saying, presupposing and implicating: How pragmatics modulates commitment. *Journal of Pragmatics*, *133*, 15–27.

Meibauer, J. (2005). Lying and falsely implicating. *Journal of Pragmatics*, *37*(9), 1373–1399.

Meibauer, J. (2014). *Lying at the semantics-pragmatics interface*. Berlin: Mouton de Gruyter.

Meijering, B., van Rijn, H., Taatgen, N. A., & Verbrugge, R. (2012). What eye move-ments can tell about theory of mind in a strategic game. *PLoS ONE*, *7*(9), e45961.

Mol, L., Verbrugge, R., & Hendriks, P. (2005). Learning to reason about other people’s minds. In L. Hall, D. Heylen, S. Marcella, C. Pelachaud, P. Wallis, & S. Woods (Eds.), *Proceedings of the joint symposium on virtual social agents* (pp. 191–198). Hatfield, UK: University of Hertfordshire.

Montague, R., Navarro, D., Perfors, A., Warner, R., & Shafto, P. (2011). To catch a liar: The effects of truthful and deceptive testimony on inferential learning. In L. Carlson, C. Hoelscher, & T. F. Shiple (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 1312–1317). Austin, TX: Cognitive Science Society.

Morency, P., Oswald, S., & de Saussure, L. (2008). Explicitness, implicitness and commitment attribution: A cognitive pragmatic approach. *Belgian Journal of Linguistics*, *22*(1), 197–219.

Nicolle, S., & Clark, B. (1999). Experimental pragmatics and what is said: A response to Gibbs and Moise. *Cognition*, *69*(3), 337–354.

Noveck, I. A. (2001). When children are more logical than adults. *Cognition*, *86*, 253–282.

Noveck, I., & Sperber, D. (2007). The why and how of experimental pragmatics: The case of “scalar inferences”. In N. Burton-Roberts (Ed.), *Pragmatics* (pp. 184–212). Basingstoke, UK: Palgrave Macmillan. https://doi.org/10.1057/978-1-349-73908-010

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics/pragmatics interface. *Cognition*, *86*, 253–282.

Pavan, S. (2013). Scalar implicatures and iterated admissibility. *Linguistics and Philosophy*, *36*, 261–290.

Pinker, S., Nowak, M. A., & Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of Sciences*, *105*(3), 833–838.

Pouscoulous, N., Noveck, I., Politzer, G., & Bastide, A. (2007). Processing costs and implicature development. *Language Acquisition*, *14*(4), 347–376.

Primoratz, I. (1984). Lying and the "Methods of Ethics". *International Studies in Philosophy*, *16*, 35–57.

Pryslopska, A. (2013). Implicatures in uncooperative contexts: Evidence from a visual world paradigm. Poster presented at XPRAG, Utrecht, The Netherlands.

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ransom, K., Voorspoels, W., Perfors, A., & Navarro, D. (2017). A cognitive analysis of deceptionwithout lying. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 992–997). Austin, TX: Cognitive Science Society.

Reboul, A. (2017). Is implicit communication a way to escape epistemic vigilance? In S. Assimakopoulos (Ed.), *Pragmatics at its Interfaces* (pp. 91–112). Berlin: Mouton de Gruyter.

Recanati, F. (2003). Embedded implicatures. *Philosophical Perspectives*, *17*(1), 299–332.

Recanati, F. (2004). "What is said" and the semantics/pragmatics distinction. In C. Bianchi (Ed.), *The semantics/pragmatics distinction* (pp. 27–43). Stanford, CA: CSLI Publications.

Rees, A., & Bott, L. (2018). The role of alternative salience in the derivation of scalar implicatures. *Cognition*, *176*, 1–14.

Rothschild, D. (2013). Game theory and scalar implicatures. *Philosophical Perspectives*, *27*(1), 438–478.

Stahl, D. O., & Wilson, P. W. (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, *10*, 218–254.

Stiller, A., Goodman, N., & Frank, M. (2011). Ad-hoc scalar implicature in adults and children. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 2134–2139). Austin, TX: Cognitive Science Society.

van Rooij, R., & Schulz, K. (2006). Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation. *Linguistics and Philosophy*, *29*, 205–250.

van Tiel, B., van Miltenburg, E., Zevakhina, N., & Geurts, B. (2016). Scalar diversity. *Journal of Semantics*, *33*(1), 107–135.

Verbrugge, R., & Mol, L. (2008). Learning to apply theory of mind. *Journal of Logic, Language and Information*, *17*, 489–511.

Weissman, B., & Terkourafi, M. (2019). Are false implicatures lies? An empirical investigation. *Mind & Language*, *34*, 221–246.