Against PCI-GCI uniformity: Evidence from deceptive language in German and Chinese

Abstract

The discussion on whether some conversational implicatures (CIs) are more 'default' than the other has taken place for a long time. Relevance Theory (RT) claims that all implicatures are context-dependent and cognitively effortful, while neo-Griceans (NG) insist on the distinction between generalized and particularized CIs which are said to differ along numerous dimensions. So far, most studies focusing on computational speed showed any enrichment is more costly than the literal understanding, and therefore favor RT. In this study, a novel approach – deceptive language with false implicatures – was used to test speakers of German and Mandarin Chinese. The main findings show that (i) false GCIs correspond to lies, while PCIs are congruent with deceptions. This observation challenges RT, while it can easily be explained under NG. (ii) The response behavior of German and Chinese participants seems to be very similar when socio-cultural factors are controlled for, suggesting that this pattern is mandated linguistically. Furthermore, two control experiments reveal that the patterns observed are not due to moral judgments but that they were caused by a genuine linguistic distinction.

1. Theoretical Background

1.1 Particularized and Generalized Conversational Implicature

Inferred meanings are abundant in natural language and ever since Grice (1989), there exists a field surrounding such non-literal contributions to utterance meaning. However, a fair bit of controversy is associated with the phenomenon called implicature — and others, though we will not be concerned with them here. For illustration, consider the classic example from Levinson (2000):

(1)   a.   A: Where's John?
                       B: Some of the guests are already leaving.
                       (i) +> John might have left with the other guests.
                       (ii) +> Not all of the guests are already leaving.
     b.   A: What time is it?
                         B: Some of the guests are already leaving.
                       (i) +> It must be late.
                       (ii) +> Not all of the guests are already leaving.

As should be evident from these exchanges, implicatures, irrespective of (sometimes disputed) sub-varieties, make apparent that natural language meaning does not only feature strict form-meaning mapping akin to translation, but also relies heavily upon extralinguistic processing considerations that take into account various functional aspects concerned with language use. To return to our example, note that, despite the homomorphism of B's utterances in (1ab), the intended meaning, namely the respective answer to the questions posed by A (1a.(i) and 1b.(i)), seems to change (as indicated after the +> symbol we adopt for making explicit the implicature

throughout this article). Here, the meaning is inferred on the basis of context and only by virtue of inference do the exchanges appear coherent, a property captured by Grice's Maxim of Relation whereby the connection between both utterances is retrieved and the implicature is computed.

This costly as well as slow computation of the so-called particularized conversational implicatures (PCIs) is rivalled, some argue, by yet another inference involved in the understanding of B's utterances in (1): generalized conversational implicatures (GCIs) (1a.(ii) and 1b.(ii)). In contrast to the context sensitive nature of PCIs, GCIs, according to neo-Gricean pragmaticists (Levinson 2000, Gazdar 1979, Horn 1989, Atlas 1980), constitute a fast default inference or preferred interpretation grounded in lexical semantics and, thereby, are always derived unless the context suggests otherwise. In other words, in both of our example sentences (1a) and (1b), the mere use of *some* would trigger the GCI +> *not all*, irrespective of the vastly different contextual conditions. In Grice's framework, this would involve the Quantity Maxim, in Horn's terminology the inference is based on the Q-Heuristic. Scalar implicatures, named as such because of their informativity ranking together with other lexemes[1] – see the Horn (1972) scales in (2) – are a paradigm case of neo-Gricean pragmatics. With the ordering proceeding from less to more informative, the neo-Gricean view propagates inferred negation of the stronger alternatives upon utterance of lower-ranking lexemes.

(2)   a.   ⟨ some, all ⟩
        b.   ⟨ possible, certain ⟩

Thus, while PCIs are costly and strongly context dependent inferences, GCIs are linguistically triggered default inferences that arise unless the context suggests otherwise. In contrast to the frameworks discussed until now, Relevance Theory advocates for a different mechanism underlying the process of inference retrieval (and meaning in general, as sentence meaning on this view is not understood to be propositional). Utterance content is said to merely serve as an indication for the speaker's meaning, and crucially, not to encode it. Instead of Gricean Maxims or their descendants, two Principles of Relevance are invoked to satisfy the hearer's expectation of relevance of any utterance directed his way. In effect, these principles, in contrast to the origin in rationality that is maintained by Griceans and neo-Griceans alike, are heuristics that follow from general cognition and enable efficient communication. These descriptive principles are explicated in the following (Sperber & Wilson 1995:260–272):

(3)   a.   **Cognitive Principle of Relevance**
                Human cognition tends to be geared to the maximization of relevance.
        b.   **Communicative Principle of Relevance**
                Every act of overt communication conveys a presumption of its own optimal relevance.

On the basis of these, relevance theorists argue, the communicative process becomes predictable, and deficient (that is, underspecified) meanings, which indicate the speaker's conversational intent and contribution, can be accounted for by employing the heuristics until an interpretation meeting the hearer's expectation of relevance is reached.

This conception, in addition to the dissociation from lexical semantics, induces a rejection of the PCI/GCI distinction since all inferences as well as the enrichments of the explicit content function in a uniform manner; in turn, nothing sets the particularized and the generalized

---

[1] In more formal terms, the ordering is construed via asymmetric entailment relations, such that stronger scalar terms asymmetrically entail the weaker ones.

enrichments in (1a) and (1b) apart (neither cognitively nor theoretically, Noveck & Sperber 2007: 11). There is, however, room for variation between GCIs and PCIs within RT, albeit not strictly linguistic in nature, such that every communicative act comes with its own assumption of optimal relevance, which may vary based on communicative circumstances and the type of utterance (Sperber & Wilson 1995: 266–78):

(4)   **Presumption of optimal relevance**
     a.   The utterance is relevant enough to be worth processing.
     b.   It is the most relevant one compatible with the communicator's abilities and preferences.

In (4b), the addition of preference allows for extralinguistic factors to affect implicature calculation, such as the cultural setting of a linguistic community or the properties of the present communicative situation. With GCIs and PCIs specifically, one such preference might be the pervasive enrichment via GCIs, which apply in wide range of contexts, over the more restricted distribution of PCIs. It is worth reiterating, however, that all observable differences between the two are assumed to occur extralinguistically, as the underlying mechanism that gives rise to both inferences is the same.

Over the last couple of decades, there have been a number of experimental investigations seeking to elucidate the ongoing debate between neo-Griceans and the post-Gricean Relevance Theory with experimental findings.

Up until now, investigations intent on providing empirical support for any of the theories just reviewed focused mainly on language acquisition and the processing implications of both major theoretical approaches to pragmatics. These endeavors, mainly operationalizing different predictions that follow from the distinct theoretical positions favored, created an extraordinarily large body of literature. The relevant predictions are reviewed in the following. Neo-Griceans, claiming that, due to efficiency considerations, GCIs are default inferences that feature both a close connection to lexical semantics and relative invariance as regards contextual parameters, predict inhomogeneity between costly PCIs and less costly GCIs. The cancellation of the latter enrichment — then the literal meaning —, on the other hand, is presumed to increase processing exertion, yielding a slower inferential process. Contrastingly, according to Relevance Theory, all enrichments follow the same principle: the paramount goal of achieving optimal relevance of the perceived input. By extension, the literal meaning, if any such concept is even adequate on this view, should be computed faster than inferences at the implicit level; further, as the PCI/GCI distinction is disregarded, both should fare similarly: slower than the retrieval of explicit meaning, with implicature cancellation playing no role at all.

Studies on whether Relevance Theory (RT) or neo-Gricean (NG) make the correct predictions focused mainly on the computational factor of speed of interpretation, as described above. In measuring reaction times, experiments seem to favor RT — prominent here are Bott & Noveck (2004), who show that the enriched interpretation of the scalar term *some* (*some but not all*) takes longer than the literal one. This slowdown supports the relevance-theoretic account which predicts that the enriched meaning of a scalar term requires more processing efforts. Similarly, a study done by Breheny et al. (2006) on disjunctive phrases concludes that the enriched interpretation of the scalar term *or* is more costly (longer reading times), consistent with RT. Developmental studies such as Noveck (2001) or, more recently Huang & Snedeker (2009), too, support the idea espoused by RT that the non-literal meaning of scalar terms requires more processing efforts than the literal one and is thus less accessible to children than to adults.

Additionally, this idea is also supported by experimental studies drawing comparisons between native and non-native speakers. Chen (2019) for instance shows that nonnative speakers are significantly more inclined to accept logically correct but pragmatically infelicitous sentences with *some* and *can* in German than native speakers, indicating that the scalar interpretation requires more processing cost than the literal one.

More recently, Huang & Snedeker (2018) conclude a series of experiments by stating that scalar implicatures always incur a processing cost, but may sometimes appear on par with literal interpretations. This is, they posit, due to predictive computation: an ideal participant may anticipate the requirement of repeated inference computation and, because of this prediction, preemptively compute the implicature, disguising the processing effort.

As of yet then, and at least under Huang and Snedeker's interpretation, experimental evidence favors Relevance Theory. However, more and more results challenge this interpretation. Inconsistent results are reported regarding the processing signature of scalar implicatures: while some find implicatures immediately available (e.g., Grodner et al. 2010; Breheny et al. 2013), others find that they do incur a cost (e.g., Bott & Noveck 2004; Cremers & Chemla 2014, Panizza et al. 2009, Huang & Snedeker 2009, 2011). Degen & Tanenhaus (2016), who found both instantly available implicatures and delayed ones in a series of experiments, argue that the largely inconsistent time-course of scalar inferences is determined by confounding factors such as the presence of lexical competition, mainly numerals, and set sizes. De Carvalho et al. (2016) showed that the presentation of a scalar term that is needed for the interpretation of a scale mate induces a priming effect for word recognition. In their experiment, *all*, which factors into the interpretation of *some* (either *some and possibly all* or *some but not all*), was shown to enhance reaction times, but not vice versa. Under a NG construal, this is expected due to the assumption of lexical scales, while with the general mechanism assumed to underlie all implicatures in RT, it is not. This provides evidence for the fact that, at least in the realm of scalar GCIs, scalarity does play a role.[2] Thus, the debate around implicature theory is still very much alive.

A further focal point of this investigation concerns the question whether the PCI/GCI distinction, if tenable at all, is a universal of human language processing. For although the universality of pragmatic principles (Grice's Maxims, Levinson's heuristics) is often implicitly taken for granted — and in the case of RT explicitly presupposed —, there are approaches that argue against this assumption. In particular, it might be possible to conjecture that the set of Maxims that govern pragmatics is not determined by the cognitive and linguistic system alone, but that it is additionally influenced by extralinguistic factors in the surrounding speech community of any given speaker. Instrumental for the formation of that view, Keenan (1976) observed that, different from Western practice, Malagasy speakers normally do not disclose the person, time or place in conversation even if they know them, and, therefore, seem to have different conversational expectations, which could result in potentially different conversational maxims, and possibly implicatures. Keenan suggested that the "Gricean analysis retains usefulness but within the perspective of a comparative typology in which locally valid systems may differ strikingly in what is marked and unmarked. An ethnographic base and ethnological comparison are required." (p.67) However, most studies that investigated this aspect concentrated on a limited subset of languages (mainly German, English, French) and left non Indo-European languages explored only sparsely, if at all.

---

[2] Note that this leaves untouched the question of processing, the investigation of which favors RT. We refer the reader to the original studies for discussion on this point. We will take up this point again in the General Discussion.

1.2 Deceptive language as a methodology in implicature theory

Recently, the emergence of a new interest area has provided a testing ground for the discussion at hand, along with others. With the rise within pragmatics of studies aiming at deceptive language use, we propose a novel way of distinguishing between the theoretical predictions and the alleged difference between PCIs and GCIs: their suitability for lying, as opposed to merely deceiving. In a recent investigation concerning language acquisition, Antomo et al. (2018) demonstrated that deceptive language motivates children to both compute the relevant implicature and take them into account in an Acceptability Judgment Task, which runs counter to earlier studies (see Katsos & Bishop 2011 for discussion).

In order to make explicit the predictions alongside their reasoning, let us first introduce the relevant categories that were developed in the study of deception. The notion of deception (in the broad sense) usually serves as a hypernym for a multitude of dishonest acts, ranging from lying to misleading, each divisible into numerous subtypes; one such category, also named deception (in the narrow sense) is defined as a "move or action … which has a goal of getting the hearer to make a false assumption" (Vincent & Castelfranchi 1981: 753). To lie, on the other hand, following Mahon (2008), is "to make a believed-false statement to another person with the intention that the other person believe that statement to be true". The striking disparity consists in the verbal component: when absent, for example when staying silent or redirecting road signs, the act in question is not, so the consensus seems to be, a lie. This notion also seems to be borne out in Mandarin Chinese when we consider the expression employed for lying in opposition to the one for deception: in the Mandarin expression for "to lie", i.e. 谎言 (*huangyan*, 'lie speech'), the first character 谎 has the radical 讠 which stands for (spoken) words or speech and the second character 言 is speech itself. Here, the connection to the construals in the recent literature seems obvious: lying, even in lexical terms, is characterized by a focus on verbal acts. In contrast, the Mandarin expression for deception, i.e. 欺骗 (*qipian*), does not employ any component relating to verbal acts. We will come back to Mandarin later (Experiment 2).

Though without immediate connection to the theoretical distinctions, a conclusion might be that deception, by way of indeterminate lexical encoding in Mandarin Chinese, is generally felt to be a broader category that subsumes both verbal and non-verbal acts.

These terminological as well as theoretical distinctions gave rise to a fair bit of controversy regarding implicatures employed to deceive, the most prominent of which we will present now. Adler (1997) categorizes untruthful implicatures, regardless of their neo-Gricean classification, as 'mere' deception, more akin to non-verbal than spoken acts – Horn (2017a, 2017b) and Saul (2012), who moreover emphasize that lying is a semantic and not a pragmatic phenomenon, assume a similar position regarding the classification of untruthful implicatures. Meibauer's (2005: 1384) point of view differs drastically: "The crucial point is that a false implicature only comes about through a verbal act of assertion to which it is bound. In contrast, a characterization of false implicature as deception misses this connection and remains purely terminological." Sweetser (1987: 60) argues that untruthful GCIs are more closely related to prototypical cases of lying than PCIs because of the former's close association with linguistic form and, by extension, with the verbal statement itself.

This theoretical dispute is matched empirically where both the findings and the corresponding interpretations differ. In particular, while Weissman & Terkourafi (2019) conclude that only repetitions and cardinal numbers (the latter of which have recently been argued not to constitute

implicatures at all; see Huang, Spelke & Snedeker 2013 and much other work) are suitable as lies proper, Willemsen & Wiegmann (2017) observe participants to indeed rank untruthful implicatures as being on par with untruthful assertions. A third empirical investigation (Antomo et al. 2018) found that false implicatures are suitable for deceptive acts, but also that participants differentiate between what is said and what is implicated. Second, the contrast between GCIs and PCIs only featured in truthful scenarios, whereas untruthful GCIs and PCIs were shown to be judged alike. In that study, there were no separate categories for lying and deception, however, so that the results do not bear on the decision between the different approaches to classification.

In a similar fashion, Weissman & Terkourafi (2019) asked participants to judge whether utterances triggering an untruthful implicature were lies on a 7-point scale ranging from "definitely not a lie" to "definitely a lie". Their results show a lot of variance between the different types of implicatures (spanning both 11 types of GCIs, across all the heuristics in the Levinsonian framework, as well as 4 PCIs), the general pattern being, however, that most items were rated in the space between "definitely not a lie" and the midpoint "neither a lie nor not a lie". Despite the obvious problem posed by the considerable variation between the implicatures and the overall trend, the conclusion that inferences tend towards not being lies is not without challenge. In particular, the study faces the ubiquitous problem of lacking a "deception" category, as with Antomo et al. (2018). On account of this, participants potentially had to resort a rescue strategy, which potentially explains the rating cluster near the middle of the scale for some of the items. Further, it has to be noted that each type was only instantiated by one item.

A further recent study carried out by Franke et al. (2019) featured a signaling game in which participants had to describe cards to a virtual co-player, depending on the assigned condition either truthfully or with the intention to make the other player choose the wrong card. While a portion of the descriptions featured only simple assertions, others made use of numeral expressions, scalar (*some*), and *ad hoc* implicatures. In this experiment, participants used implicatures to deceive, but in a less consistent fashion than with literal meaning; moreover, the three different kinds of triggers were not exploited equally. In this sense, Franke et al. observed a scale across their investigated triggers, on which numerals feature close to truth-conditional content, *ad hoc* implicatures appear very far from it, and *some* occurs somewhere in between the two.

In sum, for the present experiments, we assume that lies are bound to verbal acts of communication, while deception in the narrow sense is applicable to non-verbal acts. This, coupled with the earlier discussion of PCI versus GCIs, yields the following predictions: following the neo-Gricean conception of pragmatics, PCIs and GCIs will not behave uniformly. This is mainly due to the fact that PCIs are less bound to linguistic form and highly context-dependent, which contrasts with contextually stable GCIs, which furthermore feature a close link to lexical semantics mediated by (the assumption of) lexically codified scales – at least in the context of scalar GCIs, which are the only type we will be concerned with here; essentially, they are understood to be more linguistic, and thus verbal under the understanding assumed here, than PCIs. However, PCIs and GCIs might be subject to different extralinguistic factors that influence the two to a different degree. Then, however, we would expect for any difference to be potentially instable cross-linguistically. As the general mechanism underlying the two is the same, any inhomogeneity should be mediated by socio-cultural factors in a deterministic fashion.

Regarding the literature on deception, Meibauer (and Adler), who highlights the close tie of the implicature to its triggering utterance, predicts that untruthful implicatures as a uniform class

constitute lies proper, while the opposing position, comprised of Saul, Horn, and others, predicts that speakers will judge implicated deceptive acts, by virtue of their non-verbal nature, as deceptions.

In order to test whether speakers of unrelated languages and from differing cultural backgrounds display comparable behavior in pragmatics, we will test both speakers of German and of Mandarin Chinese in the realm of deceptive implicatures.

Supposing that speakers employ a single mechanism to derive inferential meaning components, as in RT, it is to be expected that GCIs and PCIs should pattern in similar ways and that any differences that do occur have to be mediated by extralinguistic factors. As these will be controlled for in the following series of experiments, we expect that speakers of Mandarin Chinese and speakers of German should display the same response behavior.

Under a neo-Gricean conception of pragmatics, on the other hand, socio-cultural differences should play a minor role, at most. That is, since PCIs and GCIs only share their status as inferences, but are assumed to occur for vastly different reasons and by distinct machinery, we expect any extralinguistic factors to be dwarfed, yielding a uniform behavior between the members of the two linguistic communities that were under investigation here, while GCIs and PCIs are not expected to elicit similar judgments due to their underlying distinction in that framework.

Barring some socio-cultural factors (see the Materials section in 2.1), we will assume that deceptive language is universal and therefore both groups will behave essentially the same in the contexts under investigation and that any culture specific deviations are limited to varieties of deception more closely intertwined with cultural identity, like prosocial and blue lies as well as humor (see Lee 2013 for discussion) – none of which were considered here.

2. Experiments

2.1 Experiment I: German
In the first experiment, we assessed whether PCIs in comparison to GCIs would be rated more as deceptions and, consequently, are regarded as non-verbal acts. Therefore, we conducted an online rating experiment using the open source software OnExp.

Participants
For the first experiment, we tested 43 native German speakers (34 female and 9 male). All participants were students from the University of Göttingen between 18 and 40 years of age (mean = 23.51 ± 4.34 years).

Material
Our online questionnaire consisted of 12 critical items (comprised of 6 untruthful PCIs and 6 untruthful GCIs) and 24 filler items (6 truthful assertions, 6 truthful PCIs, 6 truthful GCIs and 6 deceptions). The items were presented as short stories. Each story involved two characters (children and/or adults) and ended with a question-answer pair, except for the deceptions, as they featured non-verbal behavior only and thus did not include any dialogue. For all items we chose scenarios that could be accurately translated into Mandarin Chinese and tried to keep cultural influences to a minimum, hence excluding e.g. prosocial lies (for a general overview, see Lee 2013; for details regarding our experiment, see Material Exp. 2). We further disregarded the subclass of deceptive language called *misleading* which, following Meibauer (2014) and

Saul (2012), is applicable to inadvertent deception and therefore does not play a role in the present study, as we were concerned with intentional deception exclusively.

Regarding the critical items, the given answer always gave rise to an implicature (PCI or GCI) which was both untruthful as well as false; what was literally said, on the other hand, was true in all cases. With respect to false PCIs, we only tested relevance implicatures because quality implicatures (e.g. irony, metaphors) would have been too complex and cultural differences between German and Mandarin Chinese speakers were to be expected. Regarding false GCIs, we only included scalar inferences referring to lexical scales (no M- and I-implicatures according to Levinson), as we wanted to keep the items as homogenous as possible; expressions used were *some* (2x), *alright*, *warm*, *sometimes*, *to try*.[3] Doran et al. (2012) tasked participants with judging whether an utterance in a given context is true or false literally or overall, using naïve notions of literal and non-literal meaning, depending on the condition. For the eleven subtypes of GCIs, there was no consistent pattern as to whether they were understood to form part of the literal or the non-literal meaning. In order to cover a wider range, we employed multiple triggers for Q-based implicatures. The deceptions served as our baseline.

Besides the cultural transferability of all items, we made sure that the output of the untruthful items and the deceptions were as similar as possible: in each of these items there was one victim of the lie/deception and the lying/deceiving person benefited him or herself from his or her behavior, had a clear motive for the lie/deception, but did not commit any criminal act.

A translated example of an untruthful GCI, an untruthful PCI and a deception is provided here for illustration:

(5)   Untruthful GCI:
        Paul's mother wants to bake a cake, for which she needs a leavened dough. Because she still has to buy sugar, she asks Paul to mix the dough for her. Before she goes to the supermarket, she reminds him that he has to take warm but not hot water for the dough to make sure that it rises. Because Paul's favorite television series starts in a few moments, he does not pay attention to the temperature of the water and accidentally takes hot water for the dough. Although he notices the hot steam, he stirs the dough quickly and runs back to the television because he doesn't want to miss the start of the series. Later Paul's mother sees that the dough did not rise and asks Paul: "Paul, what kind of water did you take for the dough?" Paul answers: "I took warm water."
        (+> "I did not take hot water.")

(6)   Untruthful PCI
        Every morning and evening Paul's mother or father brushes Paul's teeth, which Paul does not like. Now it's late in the evening and Paul has to go to bed. With his father, he goes into the bathroom. Immediately before his father starts to brush Paul's teeth, the phone rings. Paul's father leaves the bathroom to answer the phone. In the meantime, Paul goes into his room, where his mother is waiting for him. She wants to put him into his bed and asks: "Paul, did your father already brush your teeth?" Paul answers: "We were just in the bathroom."
        (+> "In the bathroom, my father brushed my teeth.")

(7)   Deception
        Peter is participating in a race at the park. The route marked in yellow circles the outside of the park once. Before long, the other participants outrun Peter and he is in last place. Then he spots a narrow trail which passes through the park diagonally. Peter looks

---

[3] We excluded numerals as there is considerable evidence that they are semantically 'exact' (Huang & Spelke & Snedeker 2013). Furthermore, in a previous study on deceptive language use, Franke et al. (2019) found that they pattern with literal lies.

around and when he is sure that nobody is watching, takes the trail as a shortcut. Shortly before the goal, the trail flows into the yellow-marked route and Peter wins the race.

Procedure

The experiment proceeded as follows: each participant first read the stories. At the end of each story, he or she should decide in a forced choice task whether the character's behavior was to be classified as "rather a deception", "rather a lie", or "rather the truth".[4] We decided to include "rather" here as we wanted to present our rating categories as a continuum, thus encouraging participants to reveal their tendencies even though they might not be perfectly clear. At the end of the survey, which lasted from 20 to 30 minutes, an optional explanation for the participant's individual rating behavior (deception vs. lie) was requested.

Since neither the term *lie* nor *deception* were defined, one might wonder whether we can be sure that participants were aware of the differences between the two. However, our ultimate goal of the study was not to delineate between deceptions and lies. Rather, we employed deceptive language as a methodology to identify potential discrepancy between GCIs and PCIs. Furthermore, we felt that the opposition of the two response options was enough to encourage participants to discriminate between the two in a systematic way. At any rate, even if the distinction was unclear to the participants, we expected this not to be compounded by the kind of implicature, but to lead to indiscriminate response behavior with both GCIs and PCIs.

A note on disregarding "misleading", a term introduced by Saul (2012) to capture inferred lies, as an option: as the non-technical term, especially for participants, might be fairly difficult to distinguish from the deception option, we felt that its inclusion would lead to an increase in uncertainty in the study – unless a definition for all options is provided, which we decided against in order to keep all participants naïve and not to intervene too much.

Not only would the addition of misleading have had an impact on the response behavior, it would also be difficult to interpret come data analysis, since a genuine categorical difference would have been indistinguishable from a mere conflation of terms.

Design and Predictions

The study featured a 2-level within subject, between item design with the factor IMPLICATURE (GCI vs. PCI). We chose a between-items design because the context dependency of the implicatures did not allow for a reuse of the same context for both implicature types. We assumed that PCIs were rated more often as deceptions, whereas GCIs should be assessed usually as lies. In contrast to studies which support RT, in a previous study (Antomo et al. 2018) we could show a difference between true GCIs and PCIs in the assessment of their truth value, which supports a neo-Gricean approach. However, in our previous study participants could not choose between lies and deceptions.

Our assumption was based on the fact that GCIs are lexically based and, thus, do not depend on the context of the utterance, while PCIs do. Therefore, GCIs should be rated more like verbal acts, like lying, while PCIs should be regarded more as non-verbal acts, like deceptions.

Data Analysis

For the statistical analysis, we conducted a generalized linear mixed model with by-subject and by-item random intercepts only for the factors IMPLICATURE (GCI vs. PCI)[5]. Although it can lead to alpha error inflation, we had to discard the maximal model with by-subject and item

---

[4] For better readability, in the results sections, we will refer to all options as "lie", "deception", and "truth" even though the participants had to decide if the target sentence was a "rather a lie", "rather a deception", or "rather the truth".
[5] Y ~ IMPLICATURE + (1 | Subject) + (1 | Item)

random intercepts as well as by-subject random slopes because the maximal model did not converge. Therefore, we dropped the by-subject random slopes as was suggested by Barr et al. (2013).

The generalized linear regression was calculated using the statistical software R (version 3.3.1, R Core Team, 2016) by the *glmer* function from the *lme4* package (Bates et al. 2015, version 1.1-12). The option *family* was set to *"binomial"*.

Results

The results of the first experiment showed the expected chiasmatic pattern. GCIs were rated 69 times as deceptions (mean = .27, sd = .44), 161 times as lies (mean = .62, sd = .49), and 28 times as truth (mean = .11, sd = .31), whereas PCIs were rated 167 times as deceptions (mean = .65, sd = .48), 67 times as lies (mean = .26, sd = .44), and 24 times as neither lie, nor deception (mean = .09, sd = .29). Furthermore, our results show that the rating pattern of deceptions, our baseline, is similar to that of PCIs. Deceptions were rated 185 times as deceptions (mean = .72, sd = .45), 71 times as lies (mean = .28, sd = .45) and twice as truth (mean = .01, sd = .09). The absolute frequencies as well as the means and standard deviations are summarized in Table 1. The mean ratings are displayed graphically in Figure 1.

Table 1: Absolute frequencies as well as means and standard deviations of the ratings of GCIs, PCIs, and deceptions for Experiment I.

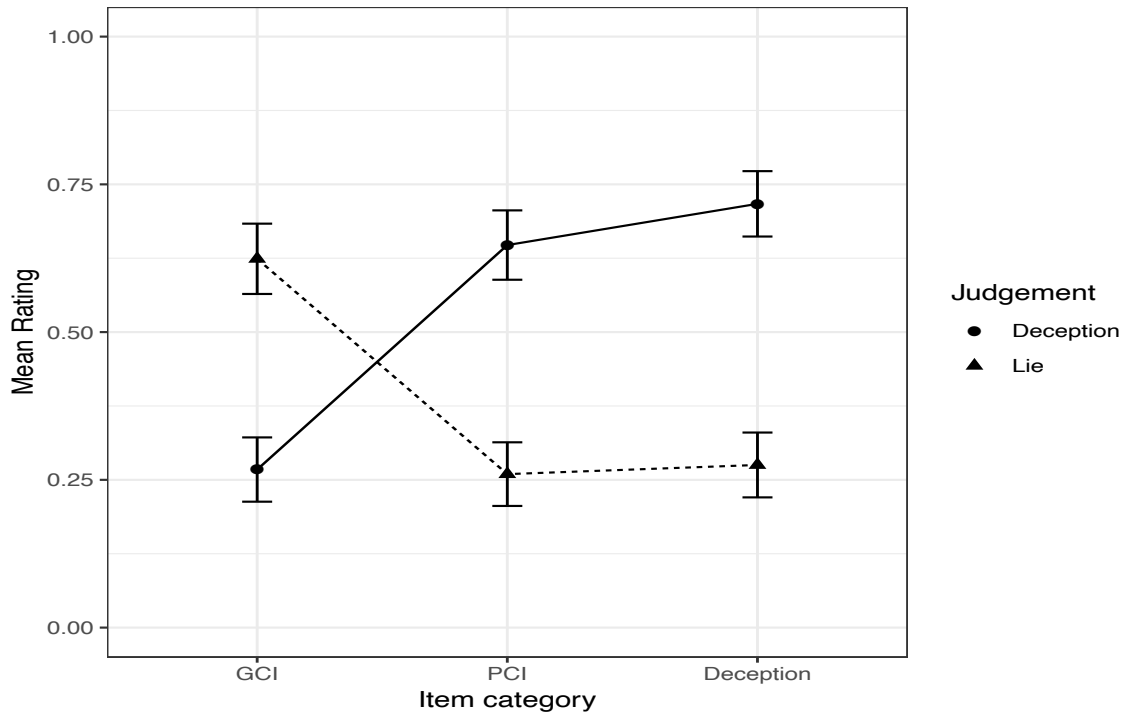|	Deception		Lie		Truth	
	Absolute Frequencies	Mean ± SD	Absolute Frequencies	Mean ± SD	Absolute Frequencies	Mean ± SD
GCIs	69	0.27 ± 0.44	161	0.62 ± 0.49	28	0.11 ± 0.31
PCIs	167	0.65 ± 0.48	67	0.26 ± 0.44	24	0.09 ± 0.29
Deceptions	185	0.72 ± 0.45	71	0.28 ± 0.45	2	0.01 ± 0.09

Figure 1: Mean ratings and SE for of GCIs and PCIs and deceptions (Experiment I)

Consistent with our assumptions, the generalized linear mixed model yielded a significant difference for the assessment of GCIs and PCIs (z = - 4.65, p < .001). We can therefore conclude that PCIs are preferably regarded as deceptions and, thus, like non-verbal acts, while GCIs are considered in general as lies. Table 2 summarizes the further results from the regression analysis.

Table 2: Results from the generalized mixed-effect linear regression for Experiment I

	Estimate	Std. Error	z Value	p value
Implicature	- 2.35	0.50	- 4.65	< .001

Furthermore, although our experiment was not designed in such a way, we compared the differences between false PCIs and deceptions, since their rating patterns were very similar (see Figure 1 and Table 1). We found no significant difference between the assessment of both deceptive behaviors (z = - 0.285, p = .78), which supports our results additionally.

2.2 Experiment II: Chinese
In the second experiment, we assessed if the results from Experiment I can be reproduced for a different linguistic community. Therefore, we reproduced the experiment in Mandarin using the same design and procedure.

Participants
For the second experiment, we tested 29 native Chinese native speakers (25 female and 4 male). All participants were students from the University Nanjing between 18 and 22 years of age (mean = 20.69 ± 0.97 years).

Material

In Experiment II, the same materials as in Experiment I were used. As we had made sure that any effects resulting from possible socio-cultural differences between both participant groups were minimized when creating our items (see Material Exp. 1), we were able to translate our materials from Experiment I without alteration – except for the protagonists' names, which were rendered into widely used and gender typical Chinese names, because the participants should not recognize that the items were translated from a Western language. In the following, we would like to shed light on the aforementioned socio-cultural factors that influenced our experiment design:

There is evidence for the fact that the concept of lying is impacted by social and cultural practices. Among children on the one hand, experiments showed that modesty-motivated untruthful statements are rated more positively by a Chinese group than by a Canadian group (Lee et al 1997, Fu et al. 2010), and the Chinese young participants rated untruthful statements that support collectivist interests more positively (Lau et al. 2012). Among adults on the other hand, Fu et al. (2001) observed that, given a prosocial context, about half of the Chinese adults did not consider untruthful telling as lies, while their Canadian counterparts categorized it as "lie" regardless of the situation. Chen et al. (2013) chose Americans and Chinese for participant groups and suggested upon their results that the element of benefit (benefit for self vs. for others) should also be added to the lying concept.

Therefore, we excluded pro-social and other-beneficial lies from both experiments in order to avoid these cultural differences and their potential effects. In our experiment, all items which were supposed to be considered lie or deception were not pro-social but beneficial for the speaker him or herself.

Additionally, the homogeneity between truth and truthfulness was also ensured in our story design. Danzinger (2010) discovered that the Americans' category of lie depended far more upon their beliefs about the utterer's belief state, while the Mopan emphasized the coincidence between the utterance and the fact. Although there is no evidence that the results from Mopan can also be observed in Chinese, truth and truthfulness in our items in Exp. 1 and 2 were designed to have the same status.

Another important issue surrounding Experiment II was the adequate translation of our questionnaire. Besides the items themselves, the three response categories required especially careful translation. Table 3 shows them in English, German and Chinese (characters and pinyin), followed by reasons and difficulties during the translation selection.

Table 3: Answer categories of the three categories used in the experiments

	lie	deception	truth
German	Lüge	Täuschung	Wahrheit
Chinese	谎言	欺骗	真话
Pinyin	huangyan	qipian	zhenhua

The term *truth* was the easiest one to translate. According to the Langenscheidt Dictionary (Ye, 2010), the word *Wahrheit* ('truth') has three interpretations (2033): the real state of affairs, the

true words, and the expression which is always true. In this case, only the second one suits our answer option, which is translated in 真话 (*zhenhua*, 'true (spoken) words').

The translation of *Lüge* ('lie') is more complicated, as it can be rendered into 谎言(huangyan, 'lie speech'), 说谎 (*shuohuang* 'talk lie') and 假话 (*jiahua* 'false (spoken) word') (1167). We chose the first translation (谎言, *huangyan*, 'lie speech') because the second one is often used as a verb, but in our original German option, it was a noun. The third one seems literally like the complement of the set *truth* "真话" (*zhenhua*, 'true' '(spoken) words'), which might have confused the participants and would have resulted in more difficulty for the translation of the word *Täuschung* ('deception').

Täuschung ('deception') was the most difficult term to translate among the three. According to Langenscheidt dictionary, it has three meanings (1773): cheating, misunderstanding and visual illusion. The last one has nothing to do with our experiment and was disregarded first. Because of the clear motivation in our design, the second one was also omitted. The first meaning offers two translations: 迷惑 (*mihuo*) and 欺骗 (*qipian*). We decided for the later one for two reasons: first, 欺骗 (*qipian*) is also given as one translation of the English word, from which the category "Täuschung" was translated. Second, 迷惑 (*mihuo*) is mostly used as a verb, whereas 欺骗 (*qipian*) can be used as both noun and verb. We also tested some synonyms of 欺骗 (*qipian*) among bilingual students, such as 蒙骗 (*mengpian*) and 忽悠 (*huyou*), and they judged 欺骗 (*qipian*) to be more suitable than the alternatives for our design.

A further difficulty arose due to the word *eher* ('rather') in the three options. Although we tried to interpret it as 更接近 (*geng jiejin*, 'closer to'), it was odd to specifiy what is "close to" the "true spoken word". However, we could not find a better alternative and as it should not affect the test, this translation was accepted.

Design and Predictions
Since we wanted to examine if the results from the first experiment are generalizable to linguistic communities other than German, we reproduced the design of the first experiment. Therefore, the second study, likewise, featured a 2-level within subject, between item design with the factor IMPLICATURE (GCI vs. PCI). We assumed that the results would show the same pattern as in Experiment I and that, thus, PCIs would generally be rated as deceptions, while GCIs would be rated as lies.

Data Analysis
In contrast to the first experiment, the maximal model did converge. Therefore, for the second experiment, we conducted a generalized linear mixed model with by-subject random intercepts and slopes, as well as by-item random intercepts for the factors IMPLICATURE (GCI vs. PCI)[6].

Results
The results of the second experiment showed the same pattern as in experiment I. In general, GCIs were primarily rated as lies (84 times as lies with a mean rating of .26 ± .44, and 46 times as deceptions with an average rating of .48 ± .50), whereas PCIs were mostly regarded as

---

[6] $Y \sim$ IMPLICATURE $+ (1 +$ IMPLICATURE $\mid$ Subject$) + (1 \mid$ Item$)$

deceptions (82 times as deceptions, mean = 0.46, sd = 0.50; 55 times as lies, mean = 0.31, sd = 0.46). In contrast to the first experiment, the Mandarin Chinese speaking participants chose the third option "rather the truth" more often. In sum, they judged untruthful GCIs 46 times (mean = .26, sd = .44)) as truth and untruthful PCIs 41 times (mean = .23, sd = .42). On the other hand, the German speaking participants chose this option on average .11 ± .31 for GCIs and .09 ± .29 for PCIs. Similar to the first experiment, the Mandarin Chinese speaking participants rated deceptions on average .76 ± .43 as deceptions (absolute frequency = 134), and .14 ± .35 as lies (absolute frequency = 25), and .076 ± .43 as truth (absolute frequency = 18). Therefore, the rating pattern for deceptions is more similar to the PCIs in comparison with the GCIs, although PCIs on average were not rated as high as deceptions or lies. The absolute frequencies and average ratings ± their standard deviations are summarized in Table 4. The mean ratings are displayed graphically in Figure 2.

Table 4: Absolute frequencies as well as means and standard deviations of the ratings of GCIs, PCIs, and deceptions for Experiment II

	Deception		Lie		Truth	
	Absolute Frequencies	Mean ± SD	Absolute Frequencies	Mean ± SD	Absolute Frequencies	Mean ± SD
GCIs	46	0.26 ± 0.44	84	0.48 ± 0.50	46	0.26 ± 0.44
PCIs	82	0.46 ± 0.50	55	0.31 ± 0.46	41	0.23 ± 0.42
Deceptions	134	0.76 ± 0.43	25	0.14 ± 0.35	18	0.10 ± 0.30

Again, the generalized linear mixed model yielded a significant difference for the assessment of GCIs and PCIs (z = - 3.811, p < .001). We can therefore conclude that in Mandarin Chinese, likewise, PCIs are primarily regarded as deceptions and, thus, like non-verbal acts, while GCIs are preferably considered as lies. Thus, our results indicate that the findings from Experiment I are not particularly bound to the speakers of German, if potential culture-specific elements are controlled. Table 5 summarizes the further results from the regression analysis.
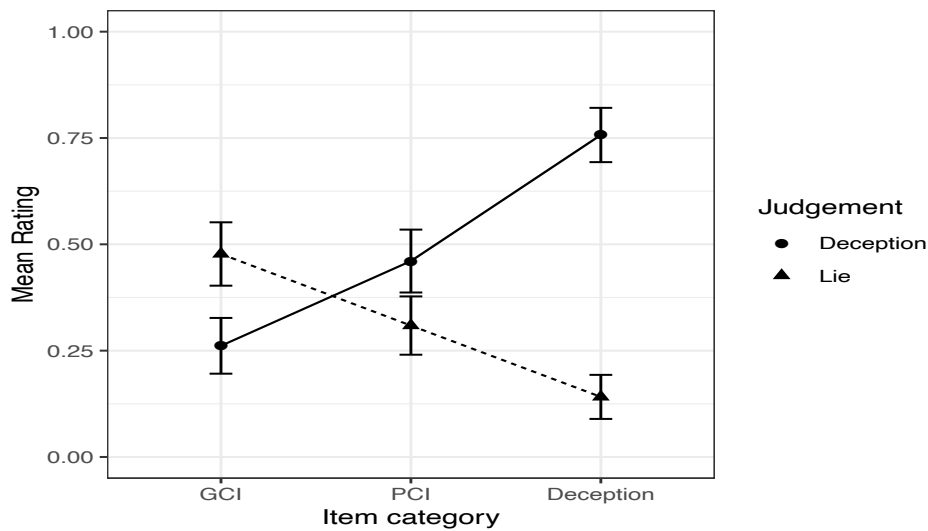

Figure 2: Mean ratings and SE for of GCIs and PCIs and deceptions (Experiment II)

Table 5: Results from the generalized mixed-effect linear regression for Experiment II

	Estimate	Std. Error	z Value	p value

Implicature	- 1.15	0.30	- 3.81	< .001

In contrast to the first experiment, by comparing the difference between false PCIs and deceptions[7] we found a significant difference between the assessment of both deceptive behaviors (z = - 4.676, p < .001). However, as mentioned above, our experiment was not designed in such a way and therefore, these results have to be examined in further studies.

2.3 Experiment III: German

In order to investigate if the chiasmatic rating pattern of untruthful GCIs and PCIs in the last two experiments was due to moral judgments, we conducted a third experiment. We wanted to ensure that participants made a linguistic distinction between lies and deceptions, i.e., that they did not consider deceptions merely as less severe lies, and, as a result, rated untruthful PCIs as deceptions, because they considered them as less immoral. This is particularly important considering the subsumption of lies as deceptions. We used the same procedure and design, but this time, we asked the participants to judge morality.

Participants

We tested 45 native German speakers between 18 and 57 years of age (mean = 22.33 ± 6.06 years). 36 of them were female, 8 were male, and one was of third gender. All participants were students from the University of Göttingen.

Material and procedure

For the third experiment we used the same items and procedure as in the first experiment, as we wanted to examine whether the observed pattern was due to participants' moral judgments. In order to obtain reliable results, we kept all parts of the experiment constant except for the participants' task to rate the morality of each item on a scale of 1 to 5 with 1 = immoral and 5 = not immoral (see section 2.1 for more information regarding material and procedure).

Furthermore, we included 6 more untruthful assertion items which served as additional filler items. Therefore, the critical items consisted of 6 untruthful PCIs and 6 untruthful GCIs, and the filler items consisted of 6 truthful assertions, 6 truthful PCIs, 6 truthful GCIs, 6 deceptions and untruthful assertions. The untruthful assertions/lies were presented in the same way as the other items. They were embedded into short stories, in which two characters (children and/or adults) were involved. Each story ended with a question-answer pair.

A translated example of a lie is provided here for illustration:

(8)   Untruthful assertion/lie:
       Charlotte and Luisa are drawing. Charlotte leaves the room for a short while to get something to drink for the two of them. Luisa prefers Charlotte's red colored pencil over her own and she continues drawing with that one, until the lead breaks. When Charlotte comes back, she asks: "What happened to my colored pencil?" Luisa responds: "The cat stepped on it."

Design and Predictions

Since we wanted to investigate if the results from the first two experiments were due to moral instead of linguistic reasoning, we used the same items as in the first two studies. Therefore, the study consisted also of a 2-level within subject, between item design with the factor

---

[7] Since the maximal generalized mixed effect model did not converge, we used a model with by-subject and by-item random intercepts only: Y ~ IMPLICATURE + (1 | Subject) + (1 | Item).

IMPLICATURES (GCI vs. PCI). In contrast to our first two experiments, the dependent variable was a rating of morality on a 5-point Likert scale. We assumed that participants were able to distinguish the linguistic difference between lies and deceptions and did not rate the items on the basis of moral judgments. Hence, we assumed that there would be no difference in moral judgments between untruthful PCIs and GCIs.

Data Analysis

To test for equivalence, we conducted a two-one-sided-t-test (TOST) procedure using the *data.TOSTpaired* function from the *TOSTER* package (Lakens 2017) in R (version 3.3.1, R Core Team, 2016). To define the equivalence bounds we used the *posterTOSTpaired* function from the same package with *alpha = 0.05*, *statistical_power = 0.9* and *N = 252*.

Since our data is ordinally scaled and the t test does typically not account for that kind of data, we additionally conducted a mixed-effect ordinal regression analysis with a Laplace approximation, by-subject random intercepts and slopes, as well as by-item random intercepts for the factor IMPLICATURE (GCI vs. PCI)[8]. The analysis was calculated using the *clmm* function from the *ordinal* package (Christensen 2015, version 2018.6-28) with the option *Hess = T*.

Results

Although the truthful items served as fillers, they were, as expected, rated higher than items featuring untruthful behavior (truthful assertions: mean[9] = 4.88, sd = 0.53, median = 5; truthful GCIs: mean = 4.59, sd = 0.82, median = 5; truthful PCIs: mean = 4.56, sd = 0.77, median = 5). Furthermore, untruthful GCIs (mean = 2.5, sd = 1.27, median = 2) and PCIs (mean = 2.55, sd = 1.12, median = 2) were rated almost identically and higher than deceptions (mean = 1.92, sd = 1.06, median = 2) and lies, which were rated as most immoral (mean = 1.57, sd = 0.95, median = 1). The results are summarized in Table 6 and are displayed graphically in Figure 3.

Table 6: Descriptive statistics (means, standard deviation and median) for Experiment III

	Mean	SD	Median
Deception	1.92	1.06	2
Lie	1.57	0.95	1
Truthful assertion	4.88	0.53	5
Truthful GCI	4.59	0.82	5
Truthful PCI	4.56	0.77	5
Untruthful GCI	2.50	1.27	2
Untruthful PCI	2.55	1.12	2

---

[8] $Y \sim Xxx + (1 + (Xxx) \mid Subject) + (1 \mid Item)$

[9] Since our data (ratings on a five-point Likert scale) are non-parametric, means are not ideal to describe our results. However, they provide a better summary to see the pattern than non-parametric characteristics like medians. Therefore, we will use means in further discussions.
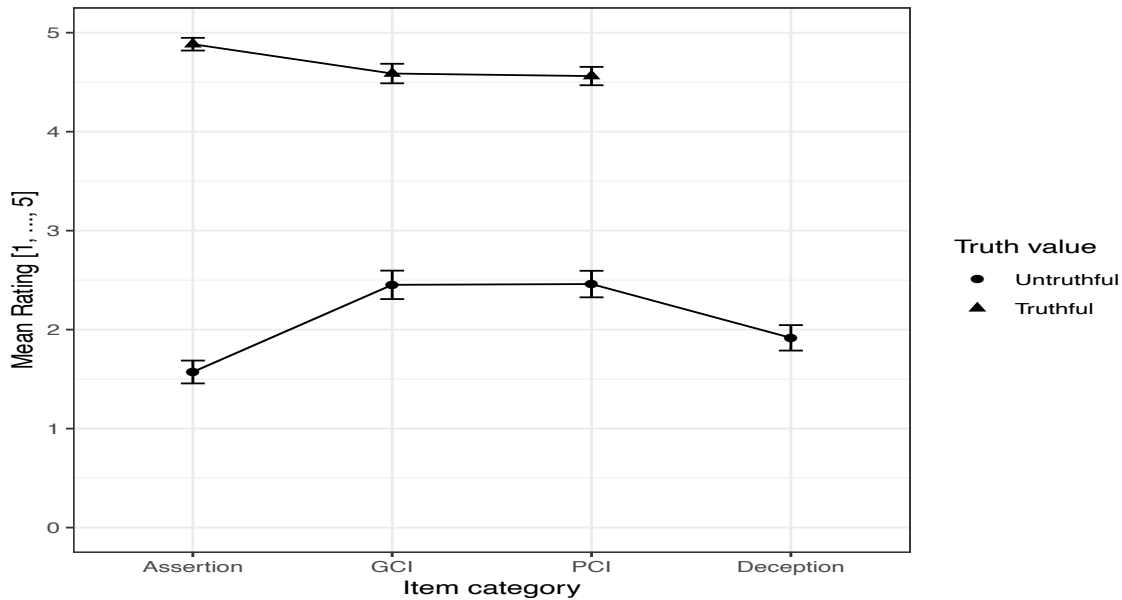
Figure 3: Mean ratings and SE for all items (Experiment III)

Consistent with our assumption, the TOST procedure yielded no difference between untruthful GCIs and PCIs (t(251) = 2.98, p = 0.002). In accordance with the TOST, the ordinal regression did not show a significant difference between both untruthful implicatures (z = - 0.115, p = 0.908, see also table 7).

Table 7: Results from the mixed-effect ordinal regression for Experiment III

	Estimate	Std. Error	z Value	p value
Implicature: PCI	0.0649	0.5639	0.115	0.908

2.4 Experiment IV: Chinese

Similar to the third experiment carried out with German-speaking participants, in the fourth experiment, we wanted to examine whether the results from the second experiment were due to moral or linguistic judgments with speakers of Mandarin Chinese. Therefore, we reproduced the items and procedure of the second experiment, as well as the design of the third experiment. Hence, for more information pertaining to material and procedure, see 2.2 Material, and Procedure, as well as 2.3 for more details regarding design and predictions.

Participants
We tested 39 participants (31 female and 7 male) between 20 and 50 years of age (mean = 26.55 ± 3.94 years). All were native Mandarin Chinese speakers from the University of Nanjing and/or the University of Göttingen.

Data Analysis
Since the statistical analysis did not differ from experiment III, see 2.3 Data Analysis for more information about the data analysis.

Results
The results from Experiment IV show the same pattern as in the third experiment: All truthful items were rated as less immoral than untruthful ones (truthful assertions: mean = 4.64, sd = 0.9, median = 5; truthful GCIs: mean = 4.73, sd = 0.65, median = 5; truthful PCIs: mean = 4.61, sd = 0.77, median = 5). Untruthful GCIs (mean = 2.85, sd = 1.41, median = 3) and PCIs (mean = 2.7, sd = 1.3, median = 2) were rated similarly, and deceptions (mean = 2.14,

sd = 1.37, median = 2) and lies were rated more immoral, with lies being rated as most immoral (mean = 1.59, sd = 1.06, median = 1). The results are summarized in Table 8 and are displayed graphically in Figure 4.

Table 8: Descriptive statistics (means, standard deviation and median) for Experiment IV

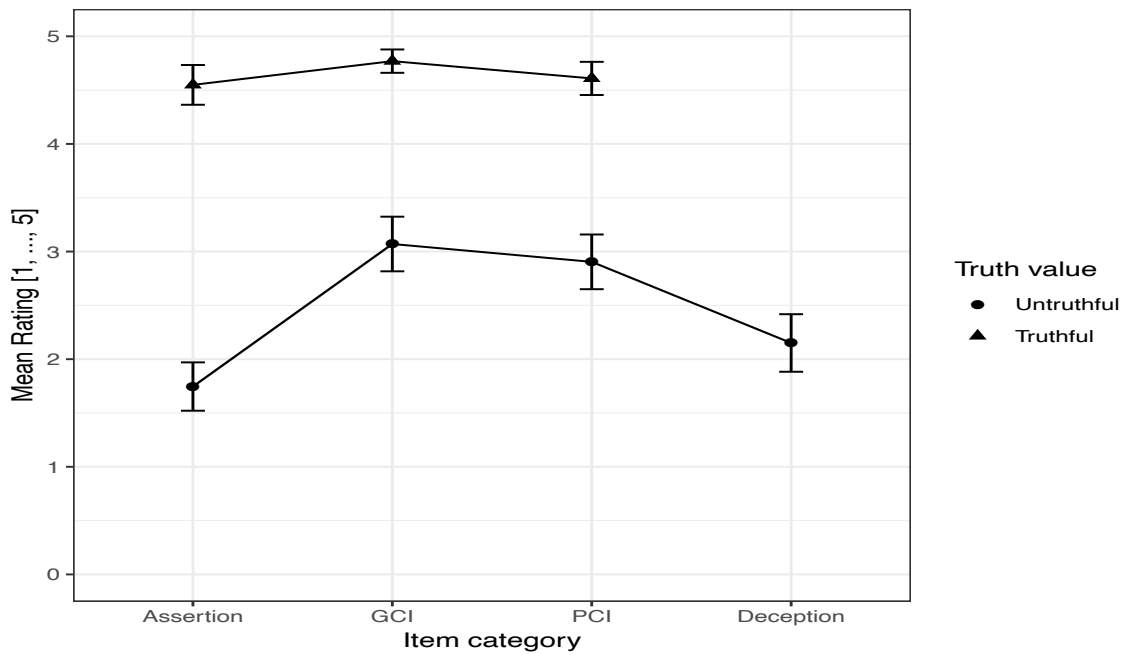|	Mean	SD	Median
Deception	2.14	1.37	2
Lie	1.59	1.06	1
Truthful assertion	4.64	0.90	5
Truthful GCI	4.73	0.65	5
Truthful PCI	4.61	0.77	5
Untruthful GCI	2.85	1.41	3
Untruthful PCI	2.70	1.30	2

Figure 4: Mean ratings and SE for all items (Experiment IV)

Consistent with our assumption and the results from Experiment III, the TOST procedure showed a significant equivalence between GCIs and PCIs (t(215) = - 1.87, p = 0.032) and the ordinal regression did not detect any significant differences (z = - 0.397, p = 0.685, see table 9). Therefore, our results do show that our Mandarin Chinese speaking participants make no difference in moral judgments of untruthful GCIs and PCIs as well.

Table 9: Results from the mixed-effect ordinal regression for Experiment III

	Estimate	Std. Error	z Value	p value
Implicature: PCI	- 0.2252	0.5549	- 0.406	0.685

3. General Discussion

We will start with our first experiment which was designed, via tasking participants with deciding between deception and lie labels for short stories involving deceptive acts, to investigate untruthful implicatures and how they relate to the pragmatics literature. First, our results, namely the significant difference between (scalar) GCIs and (relevance-based) PCIs[10], align with neo-Gricean pragmatics and seem, as far as we are aware, unexpected from a post-Gricean perspective (barring social factors determining the observed differences). What is more, they give us insight into the findings by Antomo et al. (2018), who showed that lying with implicatures is indeed possible (and is also accessible during language acquisition). Due to the employed design in Antomo et al. (2018), however, it was not possible to disentangle categorization preferences – all items were rated on a Likert-type scale and no separate category for deceptions was provided.[11]

Turning to theory first, it has yet been difficult to experimentally verify the differentiation between GCIs and PCIs postulated by neo-Griceans. In our results, however, it serves to explain the pattern we observed, such that PCIs pattern with deceptions, which are conceptualized as non-verbal acts, whereas GCIs are treated as linguistic instances (lies). Because Relevance Theory does not distinguish between different kinds of implicature (or pragmatic bottom-up processes for that matter) but instead models all as derivable from general cognitive principles, our data requires a more detailed explanation under this approach. Assuming that a single cognitive mechanism drives the generation of all inferential meanings, any difference between the two types of inference has to be relegated to the socio-cultural setting from which the implicature-triggering utterances emerge and are thus caused by extralinguistic factors, see the discussion of (4b). Note, however, that our items were designed in such a way as to minimize the impact of extralinguistic factors by using contexts that generalize well. With this in mind, it seems implausible that both speakers of German and Mandarin Chinese evolved to treating GCIs and PCIs uniformly, given that there is evidence that other types of lies are encoded and judged in vastly different ways (see Lee 2013).

There is, however, another avenue for explaining the results without resorting to a linguistic cause. One might argue that untruthful PCIs are just weak lies and that the "deception" category in our first experiments was employed by participants to mark a weaker category than "lie". On this account, the cancellability and context dependency of PCIs constitutes their being weaker than direct methods of deceiving; while GCIs, which occur in a wider range of contexts and are less easily cancellable, are considered stronger deceptions and thus warrant the "lie" category. This alternative account serves to explain our data without positing any linguistic inhomogeneity between the two kinds of inference.

In order to address this claim, we conducted two further experiments, one with German and one with Mandarin Chinese speaking participants, and tasked them to evaluate the very same items along a scale of morality. If indeed our results in the first experiments had been caused by moral

---

[10] Note that even those items which did not show values exactly matching a rather clear-cut classification as either lie or deception nevertheless conformed with the pattern described above; for them, the difference between ratings as lie/deception was simply smaller.

[11] In conjunction with our results, this strengthens the view that lying is a continuum of which deception in the narrow sense is a part. By not providing the choice of deception participants were forced to use the lying scale and had no problem in evaluating the items in these constrained conditions, even though we now have reason to believe that participants do differentiate between different levels of deception just like posited in the pragmatics literature. It would be interesting to see whether children are able to detect this discrepancy as well as the adults investigated here.

rather than linguistic factors, a similar pattern was to be expected in those later experiments as well, such that untruthful PCIs were identified as less morally aberrant than untruthful GCIs. On the contrary, however, participants rated both untruthful GCIs and PCIs in overall similar ways, while clearly differentiating between implicatures on the one hand and assertions and deceptions on the other. That is, the control experiments advocate for the interpretation that the patterns observed in the first experiments are not due to moral judgments but that they were caused by a linguistic distinction.

What is more, the fact that untruthful assertions and deceptions were judged as worse moral transgressions than GCIs and PCIs indicates that participants take directness into account when evaluating morality. Coupled with our findings that deceptions patterned with PCIs and assertions with GCIs, this shows that deceptions and lies are not simply variants of the same phenomenon in different strengths, but distinct categories to the exclusion of morality concerns. In light of this, then, the differences observed in our Forced Choice experiments cannot be explained in terms of severity, but are, in our view, caused by genuine linguistic differences between the two types of implicature, as posited in neo-Gricean pragmatics.

Furthermore, we would like to highlight a more methodological issue. Using deceptive language in experimental pragmatics, in certain contexts, is a suitable method to carve out theoretical taxonomies. In Antomo et al. (2018), it is shown that deceptive language is a good method for revealing children's pragmatic competence – in this study, the findings suggest that children as early as age 5 are able to compute implicatures, a finding that challenges earlier results such as Noveck (2001). Apparently, the use of implicated lies in order to establish a high relevance of the task for the children and to increase the at-issueness of the implicatures proves to be a setting which brings to the fore children's inferential abilities. Similarly, in our study deceptive language proves to be a suitable way to bring to the fore theoretical differences between GCIs and PCIs. As far as we are aware, the present experiment is one of the few which was able to give credence to a neo-Gricean view of pragmatics, the other, discussed in the theoretical background section, being de Carvalho et al. (2016), who provided evidence for the existence of lexical scales as employed in the theoretical treatment of a subvariety of GCIs. There is a point to be addressed here, however: most experimental findings support Relevance Theory because there seems to be no processing difference between different kinds of implicatures and because all inferences are more costly than the literal meaning, contrary to theoretical assumption (Levinson 2000); a pattern often taken to indicate that neo-Gricean pragmatics faces serious challenges empirically. Despite this apparent conundrum for the neo-Gricean camp, results about processing scalar implicatures vary wildly, up to the point of being inconclusive (see Theoretical Background section).

How does this relate to the conclusions presented above? We would like to advocate for a distinction between offline judgments and online processing of inferences (see Lohiniva & Panizza 2016; Thalmann & Panizza 2019 for similar arguments and experimental results corroborating this conclusion). Chierchia (2013), for example, remains neutral on the processing demands, while still maintaining the theoretical distinction. This position is indeed compatible with the results obtained in other studies and has explanatory value for our own, which focused exclusively on the point of distinctness arising from lexical triggering of an implicature or the absence of that feature. Whether there is actually no processing dissimilarity associated with the different implicatures or whether the homogeneity witnessed in these studies is induced for other reasons is, of course, not a question we feel prepared to answer. As a last point surrounding this issue, it is important to note that other implicatures, i.e., non-scalar GCIs and non-relevance PCIs, might produce a divergent picture. We leave these issues for further research. A tentative conclusion of this in conjunction with the results of our Experiments I and II might be that, as hinted above, language users have direct access to

theoretically established taxonomies, at least in terms of offline behavior; without direct implications for experiments measuring computational effort. More research is needed for substantiation, however.

A further point concerns Meibauer (2014), who, while working in a neo-Gricean framework, predicts that PCIs should be perceived as lies proper since they are crucially related to assertions as the explicit utterance forms the triggering environment for the inference. This proximity, he argues, should facilitate similarity to asserted lies. To reiterate our findings, we are faced with the opposite arrangement such that GCIs correspond to lies, while PCIs are congruent with deceptions.

In a similar vein, the observed pattern speaks against Horn (2017a, 2017b), who espouses the view that inferences in general are to be understood as constituting deceptions. While this lines up with the outcome for PCIs, this is not the case with GCIs; and so, a more fine-grained distinction is needed. This is not only supported by the respective choices favored by the participants when comparing GCIs and PCIs, but also based on the non-significant difference between the latter and action-based deceptions. In other words, despite the linguistic triggering context of implicatures, participants still likened them to actions like swapping the sugar shaker with the one containing salt. In addition, one of our non-implicature deception items involving fabricated credentials for a potential employee position was judged more on par with GCIs than with PCIs. Here again, participants exposed their tendency towards treating strictly linguistic acts as lies, even when the utterance which gives rise to the deceptive effect is not explicitly mentioned.

This assumption is supported by participants' own explanations of their rating behavior, which were requested at the end of the survey. Both Mandarin Chinese and German speakers roughly provided two approaches to justify their answers. On the one hand, the majority of the participants classified lies as explicitly untrue utterances whereas deceptions were described as indirect and often based on actions; on the other hand, there were participants who linked their ratings to a moral judgment of the deceptive behavior and thus explained that they chose "lie" for harmful behavior or whenever they suspected the deceiver of an evil intention, whereas "deception" was chosen for less severe instances. Notably, in light of the results for Experiments III and IV, morality or severity do not seem to play a role in distinguishing untruthful GCIs from PCIs as both were rated equally by both speakers of German and Mandarin Chinese.

Apart from the implications for the theoretical treatment of lying in particular and implicatures more generally, we also wanted to investigate the generalizability of our findings. To do this, we compared the results we obtained with German participants to a Mandarin Chinese version of Experiment 1. Mandarin Chinese is of particular interest in this respect since many studies have reported that, at least when socio-cultural factors are concerned, there is a substantial amount of disparity with Western conceptions of deceptive acts and their moral significance (for details, see Section 1). As our results show, however, for the implicatures under consideration, if the culture-specific issues are controlled for, there seems to be a very similar response behavior, and thus not even cultural idiosyncrasies supersede the general tendency of speakers to liken PCIs to deceptions, and GCIs to lies. We take this to indicate (i) that the distinction between GCI and PCI, leaving aside matters of processing, is indeed tenable (contra Breheny et al. 2006, Noveck/Bott 2004), (ii) that the two inferences are of divergent status with regard to deceptive acts, and (iii) that despite variation in social landscape, the categorization pattern is mandated by the aforementioned linguistic distinction, i.e., (non-)resemblance to verbal acts.

The only caveat worth mentioning is the general tendency of the Mandarin Chinese speakers to choose the "rather true" option with both false GCIs and PCIs more than the German participants and the resulting significant difference between PCIs and deceptions. One explanatory avenue might be to argue that, for an undetermined reason, speakers of Mandarin Chinese side with a semantic interpretation of the target utterances, whose truth value invariably was truth, rather than taking into account the implicature, while German speakers display a greater proclivity towards inferred meaning. Another reason along the same lines is that speakers of Mandarin Chinese were more disposed towards implicature cancellation than the German-speaking participants. As far as we know, however, there is neither a principled explanation for any such favoritism for semantic (or logical) interpretations/cancellation, nor is there reason to expect inhomogeneity in this regard.

There might also be – and this seems to us the likeliest solution because it does not imply any difference in interpretative strategies – an effect of other matters outside of the language processing facilities. One aspect in relation to this might be the chosen contexts. Despite our best efforts to reduce any culture specific confounds, there might be contexts we may have mistakenly judged to be universally applicable. For example, in one of our items the participant's judgment was dependent on his or her notion of cooking: in this scenario, after having bought pasta with prawns at an Italian restaurant and reheating it for dinner, a husband makes his wife believe that he has cooked the meal himself. For German participants, his behavior was – as intended – a case of deception, whereas Chinese speakers often chose "rather true" as their answer. Thus, it seems probable that for them, heating up a premade dish is an instance of cooking, while German speakers applied a narrower frame for this activity, excluding reheating in the process.

4. Conclusion

In this Forced Choice study, we provided evidence for the fact that untruthful implicatures can be employed to lie (a replication of Antomo et al. 2018). We were further able to show that, in opposition to most previous experiments investigating the GCI-PCI complex, there are reasons to uphold the theoretical distinction between the two. In particular, we argue that homogeneous processing signatures of PCIs and GCIs do not warrant the conclusion that neo-Gricean accounts of pragmatics make wrong predictions entirely or that Relevance Theory is more psychologically tenable. In the context of deceptive language, GCIs resemble verbal utterances and thus pattern with untruthful assertions whereas PCIs resemble non-linguistic deceptions based on actions as opposed to verbal triggers. As far as approaches to lying are concerned, our results are in discord with Meibauer (2014), who argues for the inverse of the pattern found in our experiment. Likewise, they contradict Horn's (2017a, 2017b) proposal that all inferences should constitute deceptions and thus resemble non-verbal deceptive acts.

Note that the complementary response behavior between GCIs and PCIs cannot be explained in terms of a moral continuum between higher-ranking lies and lower-ranking deceptions, as the equivalence of both implicature types in our control experiments on morality shows.
What is more, this finding appears to be stable cross-culturally as well as cross-linguistically, at least when German and Mandarin Chinese are compared as we did here. This justifies the deduction that the reflexes of implicatures in the realm of deception are inferable based on their theoretical status and that culture plays only a minimal role when weighed against linguistic considerations.

Abbreviations

+>   implicate
GCI   Generalized conversational implicature
PCI   Particularized conversational implicature
NG   Neo-Gricean
RT   Relevance Theory

References

Adler, Jonathan E. (1997): Lying, deceiving, or falsely implicating. In: *Journal of Philosophy* 94.9, 435–452.

Antomo, Mailin, Susanne Müller, Katharina Paul, Paluch, Markus & Maik Thalmann (2018): When children aren't more logical than adults: An empirical investigation of lying by falsely implicating. In: *Journal of Pragmatics* 138, 135–148.

Atlas, Gerald (1980): Pragmatics and logical form. In: *Journal of Pragmatics* 4.1, 1–13.

Barr, Dale J., Roger Levy, Christoph Scheepers & Tily, Harry J. (2013): Random effects structure for confirmatory hypothesis testing: Keep it maximal. In: *Journal of memory and language* 68.3, 255–278.

Bates, Douglas, Martin Mächler, Ben Bolker, & Steve Walker (2015): Fitting Linear Mixed-Effects Models Using lme4. In: *Journal of Statistical Software* 67.1, 1–48.

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

Breheny, Richard, Napoleon Katsos & John Williams (2006): Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. In: *Cognition* 100, 434–463.

Breheny et al. 2013

de Carvalho, Alex, Anne C. Reboul, Jean-Baptiste Van der Henst, Anne Cheylus & Nazir, Tatjana (2016): Scalar Implicatures: The psychological reality of scales. In: *Frontiers in Psychology* 7, 1–9.

Chen, Rong, Chunmei Hu & Lin He (2013): Lying between English and Chinese: An intercultural comparative study. In: *Intercultural Pragmatics* 10.3, 375–401.

Chen, Yuqiu (2019): An experimental study on scalar implicatures: Comparing German native speakers and Chinese learners of German. In: Van Alem, Astrid, Anastasiia Ionova & Cora Pots (eds), *Proceedings of ConSOLE XXVI*, Leiden: Leiden University Centre for Linguistics, 116–138.

Chierchia, Gennaro (2013): *Logic in grammar: Polarity, free choice, and intervention*. Oxford: Oxford University Press.

Cremers, Alexandre & Emmanuel Chemla (2014): Direct and indirect scalar implicatures share the same processing signature. In: Pistoia-Reda, Salvatore (ed), *Pragmatics, Semantics and the Case of Scalar Implicatures*. London: Palgrave Macmillan, 201–227.

Danziger, Eve (2010): On trying and lying. Cultural configurations of Grice's Maxim of Quality. In: *Intercultural Pragmatics* 7.2, 199–219.

Degen, Judith & Michael K. Tanenhaus (2016). Availability of alternatives and the processing of scalar implicatures: A visual world eye-tracking study. In: *Cognitive Science* 40, 172–201.

Doran, Ryan, Gregory Ward, Meredith Larson, Yaron McNabb & Rachel E. Baker (2012): A novel experimental paradigm for distinguishing between what is said and what is implicated. In: *Language* 88.1, 124–154.

Franke, Michael, Giulio Ducinati & Nausicaa Pouscoulous (2019): Strategies of Deception: Under-Informativity, Uninformativity, and Lies–Misleading With Different Kinds of Implicature. In: *Topics in Cognitive Science*.

Fu, Genyue, Kang Lee, Catherine A. Cameron & Fen Xu (2001): Chinese and Canadian Adults' Categorization and Evaluation of Lie- and Truth-Telling about Prosocial and Antisocial Behaviors. In: *Journal of Cross-Cultural Psychology* 32.6, 720–727.

Fu, Genyue, Megan K. Brunet, Yin Lv, Xiaopan Ding, Gail D. Heyman, Catherine A. Cameron & Kang Lee (2010): Chinese Children's Moral Evaluation of Lies and Truths— Roles of Context and Parental Individualism–Collectivism Tendencies. In: *Infant Child Development* 19.5, 498–515.

Gazdar, Gerald (1979): *Pragmatics: Presupposition, implicature, and logical form*. New York: Academic Press.

Grice, Paul (1989): *Studies in the Way of Words.* Cambridge, MA: Harvard University Press.

Grodner, Daniel, Natalie M. Klein, Kathleen M. Carbary & Tanenhaus, Michael K. (2010). Some and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. In: *Cognition* 116, 42–55.

Horn, Laurence R. (1972): *On the semantics of logical operators in English*. PhD thesis, Yale University.

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

— (2017a): What lies beyond: Untangling the web. In: Rachel Giora/Michael Haugh (eds.), Doing Pragmatics Interculturally: Cognitive, Philosophical, and Sociopragmatic Perspectives. Berlin/Boston/Munich: de Gruyter, 151–174.

— (2017b): Telling it slant: Toward a taxonomy of deception. In: Janet Giltrow/Dieter Stein (eds.), The Pragmatic Turn in Law. Inference and Interpretation in Legal Discourse. Berlin: de Gruyter, 23–56.

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

— (2011). 'Logic & Conversation' revisited: Evidence for a division between semantic and pragmatic content in real time language comprehension. In: *Language and Cognitive Processes* 26, 1161–1172.

— (2018): Some inferences still take time: Prosody, predictability, and the speed of scalar implicatures. In: *Cognitive Psychology* 102, 105–126.

Huang, Yi Ting, Elizabeth Spelke & Jesse Snedeker (2013): What exactly do numbers mean? In: *Language Learning and Development* 9, 105–129.

Katsos, Napoleon & Dorothy Bishop (2011): Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. In: *Cognition* 120, 67–81.

Keenan, Elinor O. (1976): The universality of conversational postulates. In: *Language in Society* 5.1, 67–80.

Lakens, Daniël (2017): Equivalence Tests: A Practical Primer for *t* Tests, Correlations, and Meta-Analyses. In: *Social Psychological and Personality Sciences* 8.4, 355–362.

Lau, Yat L., Catherine A.Cameron, Kai M. Chieh, Jennifer O'Leary, Genyue Fu & Lee, Kang (2012): Cultural Differences in Moral Justifications Enhance Understanding of Chinese and Canadian Children's Moral Decisions. In: *Journal of Cross-Cultural Psychology* 44.3, 461–477.

Lee, Kang, Catherine A. Cameron, Fen Xu, Genyao F. And & Julie Board (1997): Chinese and Canadian Children's Evaluations of Lying and Truth Telling: Similarities and Differences in the Context of Pro- and Antisocial Behaviors. In: *Child Development* 68.5, 924–934.

Lee, Kang (2013): Little liars: Development of verbal deception in children. In: *Child development perspectives* 7.2, 91–96.

Levinson, Stephen C. (2000): *Presumptive meanings. The theory of generalized conversational implicature*. Massachusetts: MIT Press.

Lohiniva, Karoliina & Daniele Panizza (2016): When Pragmatics Helps Syntax: An Eye Tracking Study on Scope Ambiguity Resolution in 4-to 5-Year-Old Children. In: Scott, Jennifer & Deb Waughtal (eds.), *Proceedings of 40th annual Boston University Conference on Language Development*. Somerville: Cascadilla Press, 216–228.

Mahon, James E. (2008): The Definition of Lying and Deception. In: *Stanford Encyclopedia of Philosophy* [http://plato.stanford.edu/entries/lying-definition/].

Meibauer, Jörg (2005): Lying and falsely implicating. In: *Journal of Pragmatics* 37, 1373–1399.

— (2014): *Lying at the Semantics-Pragmatics Interface*. Berlin: de Gruyter.

Noveck, Ira A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. In: *Cognition*, 78.2, 165–188.

Noveck, Ira A. & Dan Sperber (2007): *The why and how of experimental pragmatics: The case of 'scalar inferences'*. **In: Burton-Roberts, Noel (ed.), *Pragmatics*. Basingstoke: Palgrave Macmillan,** 184–212.

Panizza, Daniele, Gennaro Chierchia, Yi Ting Huang & Jesse Snedeker (2009): Relevance of polarity for the online interpretation of scalar terms. In: *Semantics and Linguistic Theory* 19, 360–378.

R Core Team (2016): R: A language and environment for statistical computing. Vienna, Austria.

Saul, Jennifer (2012): *Lying, Misleading, and what is Said*. Oxford: Oxford University Press.

Sperber, Dan & Deidre Wilson (1995): *Relevance: Communication and Cognition*. Oxford: Blackwell.

Sweetser Eve (1987): The definition of lie. In: Holland, Dorothy & Naomi Quinn (eds.), *Cultural models in language and thought*. Cambridge: Cambridge University Press, 43–66.

Thalmann, Maik & Daniele Panizza, (2019): Present to the eye, away from the mind. Dissociating online comprehension and offline judgments of indirect scalar implicatures. In: Brown, Megan M. & Brady Dailey (eds.), *Proceedings of 43rd annual Boston University Conference on Language Development*. Somerville: Cascadilla Press, 667–678.

Vincent, Jocelyne M. & Cristiano Castelfranchi (1981): On the art of deception: How to lie while saying the truth. In: Parret, Herman, Marina Sbisà & Jef Verschueren (eds.), *Possibilities and Limitations of Pragmatics: Proceedings of the Conference on Pragmatics*, Amsterdam: Benjamins, 749–777.

Weissman, Benjamin & Marina Terkourafi (2019): Are false implicatures lies? An empirical investigation. In: *Mind & Language* 34, 221–246.

Willemsen, Pascale & Alex Wiegmann (2017): How the truth can make a great lie: An empirical investigation of the folk concept of lying by falsely implicating. In: Glenn Gunzelmann et al. (eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society*. Austin: Cognitive Science Society, 3516–3621.

Ye, Bendu (ed.) (2010): *Langenscheidt Großwörterbuch Deutsch als Fremdsprache: (deutsch-chinesisch) = Lang-shi-De-Han-shuang-jie-da-cidian*. Foreign Language Teaching and Research Press.