GLOSSA psycholinguistics

Zhang, Y., & Davidson, K. (2024). Interpreting referential noun phrases in belief reports – the *de re/de dicto* competition. *Glossa Psycholinguistics*, 3(1): 13, pp. 1–34. DOI: https://doi.org/10.5070/G60111258

eScholarship University of California

Interpreting referential noun phrases in belief reports – the *de re/de dicto* competition

Yuhan Zhang, Department of Linguistics, Harvard University, US, yuhanczhang@gmail.com

Kathryn Davidson, Department of Linguistics, Harvard University, US, kathryndavidson@fas.harvard.edu

The *de re*/*de dicto* ambiguity centers on the referential and/or attributive properties of noun phrases in the scope of intentional operators, such as belief reports. For the belief report *Julie believes Elizabeth's poem will win the competition*, a *de re* reading of the embedded referential noun phrase *Elizabeth's poem* entails that the referential association between this noun phrase and the target poem is true from the perspective of the speaker, but may not be registered as such in the belief holder's (i.e. Julie's) mind. In contrast, a *de dicto* reading describes Julie's beliefs as she registers the referential association in her mind. While both the *de re* and *de dicto* readings of definite noun phrases are judged acceptable, given different supporting contexts, we show that the acceptability of *de re* readings is vulnerable to contextual and pragmatic manipulations. One such case involves a context in which a belief holder Julie holds a mistaken belief about the identity of the poem, for instance, by thinking that it was written by Nicole, while in reality it was written by Elizabeth. This mistaken identity context introduces a *de dicto* reading of a competing noun phrase *Nicole's poem* in *Julie believes Nicole's poem will win the competition*. Under this context, the speaker-oriented *de re* reading of *Elizabeth's poem* has a roughly bimodal acceptability distribution, while the *de dicto* noun phrase was overall preferred. Our study is the first that systematically lays out the empirical landscape of *de re*/*de dicto* readings of definite noun phrases and points out the vulnerability of the *de re* reading. The investigation solidifies the foundation for further theory development and endorses the practice of collecting reliable empirical judgment data for nuanced semantic phenomena.

1. Introduction

The *de re/de dicto* distinction refers to an interpretive ambiguity of noun phrases embedded in an intensional domain.[1] For example, in (1), under a *de re* reading of the noun phrase *a prince*, Aurora wants to marry a particular individual who the speaker of (1) knows to be a prince. This is, in fact, the scenario in the first part of the story of *Sleeping Beauty*, in which Aurora falls in love with a man she meets in a forest who the narrator knows to be Prince Phillip, although Aurora herself is not aware of his royal station. In this scenario, someone can truthfully describe Aurora's desires with (1). However, this sentence also has another interpretation – the *de dicto* interpretation of the noun phrase *a prince*, in which Aurora's beliefs are characterized as, basically, wanting to marry a prince, whoever might be. In the context of *Sleeping Beauty*, the same sentence under the *de dicto* interpretation is false because Aurora desires to marry only the man she met in the forest; thus, she is not in the state of desiring to have a husband that is a prince.

(1) Aurora wants to marry a prince.
   a. True under the *de re* reading in the context of the *Sleeping Beauty* story
   b. False under the *de re* reading, if Aurora did not want to marry Prince Phillip (and Prince Phillip is the only prince in the context)
   c. True under the *de dicto* reading, if Aurora were following expectations for royalty
   d. False under the *de dicto* reading in the context of the *Sleeping Beauty* story

A classic approach in formal semantics to model the *de re/de dicto* distinction is via scope ambiguity (Partee et al., 1990; Quine, 1956; von Fintel & Heim, 2011, a.o.). As represented in (2a) for the *de re* reading of (1), the existential quantifier ∃ takes a wider scope than the universal quantifier ∀ over possible worlds associated with Aurora's desires; the property of being a prince that holds of the bound variable x is evaluated to be true in the actual world w_0; the event that Aurora marries x takes place in the possible worlds w' that are compatible with what Aurora wants in the actual world. On the other hand, for the *de dicto* reading in (2b), the universal quantifier ∀ over possible worlds takes a wider scope than the existential quantifier ∃, and the "prince" property is evaluated to be true in Aurora's desire worlds w'.

(2) a. $\exists x\ [\mathit{prince}_{w_0}(x) \wedge \forall w'\ [\mathit{WANT}_{w_0}(A,w') \rightarrow \text{marry}_{w'}(A,x)]]$ (*de re*)
   b. $\forall w'\ [\mathit{WANT}_{w_0}(A,w') \rightarrow \exists x\ [\text{prince}_{w'}(x) \wedge \text{marry}_{w'}(A,x)]]$ (*de dicto*)

[1] As early as Aristotle, linguistic phenomena related to *de re/de dicto* have been observed. Yet this pair of Latin terms was not intensively applied until the Medieval period by Thomas Aquinas. The study of phenomena related to *de re/de dicto* in philosophy and linguistics can first be seen in Frege (1948), Russell (1905), and Quine (1956), even if none of them explicitly use the terms *de re/de dicto* in their writings. The current senses of *de re* and *de dicto* may deviate a little from the literal Latin meanings of the terms (*de re* 'of the thing', *de dicto* 'of what is said') (von Fintel & Heim, 2011), so it may be clearer to introduce the *de re/de dicto* distinction via contextualized examples. For more details of the nomenclature, see Keshet and Schwarz (2019).

The scope approach that differentiates the two readings in terms of scope interaction generates interesting implications. For the *de re* reading in (2a), since *prince* is evaluated outside the universal quantifier over the possible worlds associated with Aurora's desire, Aurora doesn't have to realize that *x* is a prince. The noun phrase *a prince* is just one of many possible referential terms that pick out the specific individual in the real world. For the *de dicto* reading in (2b), on the other hand, since *prince* is evaluated within the scope of *want*, the interpretation is that whoever Aurora wants to marry is a prince. Assuming Aurora only wants to marry one person, if she ever has someone who she wants to marry, that person should be a prince.

The *de re/de dicto* ambiguity has also been extensively studied in definite noun phrases with a scope analysis (Fodor, 1970; Nelson, 2019; Percus, 2000; Romoli & Sudo, 2009; von Fintel & Heim, 2011, a.o.). An example is (3), where the possessive noun phrase *your abstract* could have either a *de re* or a *de dicto* reading, given the corresponding supporting context. Specifically, the ambiguity hinges on whether the belief holder is aware of the possessive relation between the abstract and the addressee.

(3) John believes that your abstract will be accepted.

> *Evaluating context for de re*: John's belief may be about an abstract that he reviewed, but since the abstract is anonymous, he doesn't know who wrote it. He tells me about that abstract and he believes that it is sure to be accepted. I know that it is your abstract and inform you of John's opinion by saying the sentence above.
>
> *Evaluating context for de dicto*: You are a famous linguist that John is acquainted with. John knows your work very well and believes that you submitted a (unique) abstract to a linguistic conference. Given his general knowledge about this specific conference and his high opinion of your work, he believes that your abstract will be accepted, even if he doesn't know which one is your abstract or hasn't read it. He tells me his belief and I am retelling the belief to you.
>
> (von Fintel & Heim, 2011, p.157)

Following the scope solution for this basic ambiguity in simple contexts, more theoretical semantic tools have been proposed to capture the *de re/de dicto* ambiguity in syntactically different complement clauses, for different types of noun phrases, and under more sophisticated contexts (Charlow & Sharvit, 2014; Deal, 2018; Elliott, 2023; Keshet, 2008; Percus, 2000; Percus & Sauerland, 2003, a.o.). Diverging a little from, yet ultimately contributing to, this path, in this article, we utilize tools in experimental semantics (e.g. Cummins & Katsos, 2019) to investigate the contextual influence on the acceptability of *de re/de dicto* readings of definite noun phrases. The motivation behind going "experimental" is that, despite the clear availability of *de re* readings of definite noun phrases in examples like (3), other examples in the existing literature suggest that the acceptability of *de re* readings may vary in ways that suggest that features of the context may play a role (Anderson, 2013; Jaszczolt, 1997; Sudo, 2014; Zhang & Davidson, 2021). Given the field's growing interest in providing robust replicable linguistic evidence to support theoretical development (e.g. Davidson, 2020; Tonhauser & Matthewson, 2015, and

see new conference venues such as Experiments in Linguistic Meaning) as well as the limited experimental research on the topic of *de re/de dicto*, we believe it is crucial to understand more about the factors that influence in which contexts *de re* readings would be more (un)acceptable. We hope this line of research can further solidify the empirical foundation of the interpretation of noun phrases in intensional semantics.

The article is structured as follows. In Section 1.1, we present claims of an asymmetry in the acceptability of the *de re* reading compared to the *de dicto* reading for definite noun phrases, among broader observations of both interpretations generally being available for noun phrases in the scope of intensional operators. In Section 1.2, we reinforce the motivation for experimental investigation and highlight the potential contribution of this study to linguistic theories. In Section 1.3, we lay out a finer-grained categorization of the contexts permitting the *de re* reading and of the *de re/de dicto* ambiguity. We use these new categories to design well-controlled stimuli for empirical testing and causal analysis. In Section 1.4, we introduce the experiment outline. Then in Sections 2 to 4, we report the designs and results of three experiments. In Section 5, we propose potential explanations of the contextual effect on *de re* acceptability and conclude with an eye toward future work.

1.1 Diverging judgments of de re readings

In this section, we present existing literature suggesting that, depending on the contexts, the *de re* reading of noun phrases embedded in intensional domains can be less acceptable or more acceptable than a *de dicto* interpretation.

First, we see claims that argue for a preference for the *de re* reading of the *de re/de dicto* ambiguity from works in Default Semantics (Capone, 2011; Jaszczolt, 1999, 2005, 2015, a.o.). In this framework, where a main claim is that utterance meaning is jointly determined by its compositionality, the intention of interlocutors, and their cognitive biases in communication, Jaszczolt (1997) argues that the *de re* reading of definite noun phrases is the default and the most salient one, because the primary objective of communication is understanding the speaker's intention by securing the referent of the speaker's utterance in the conversational context. Since the *de re* reading highlights the referential property of noun phrases and is able to select objects in the broader conversational context, it should stand out during interpretation. This *default de re* theory not only predicts the availability of *de re* interpretations but also predicts them to be even more accessible relative to *de dicto* during communication.

In a very different domain, namely legal studies, Anderson (2013) reports that the *de re/de dicto* ambiguity has been overlooked in the interpretation of legal statutes – the emphasis on *de re* readings of statutes and the lack of attention to the *de dicto* reading have even led to puzzling judicial results.[2] For example, in the famous Enron Scandal in 2001, Enron's auditor Arthur

[2] For more legal cases where the bias towards *de re* reading influenced judicial results, see Anderson (2013).

Anderson anticipated litigation and urged its employees to destroy related financial documents. Their action stopped on the day when the U.S. Securities and Exchange Commission subpoenaed records. At that time, two federal obstruction statutes applied to document destruction. The more general statute makes it a federal offense to "corruptly…endeavor to influence, obstruct, or impede the due administration of justice."[3] The more specific one prohibits "knowingly… corruptly persuad[ing] another person…with the intent to…destroy an object…[or]…impair the object's…availability for use in an official proceeding."[4] The intentional words in these two statutes, *endeavor* and *intent*, grant each statute a *de re* and a *de dicto* interpretation, as shown in (4) and (5).

(4) *De re* interpretations
   (i) For the general statute: There is some X, which is in fact an instance of justice being administered. The defendant endeavors to influence or obstruct X.
   (ii) For the specific statute: There is a specific official proceeding Y in which the defendant intends to impair some objects' availability for use.

(5) *De dicto* interpretations
   (i) For the general statute: The defendant has the corrupt intention to influence what we describe as "the administration of justice" (Anderson, 2013, p. 28).
   (ii) For the specific statute: The defendant has the intention to initiate some kind of impairment against any possible official proceeding.

Because there was an overwhelming reliance among the judges on the *de re* interpretation of the two statutes and there was no sufficient evidence that suggests Arthur Anderson's destruction was under the knowledge of a *specific* pending proceeding, the defendant was not charged. What is worth noting is that the ruling could have been different had the *de dicto* reading been picked up. Anderson relies on children's difficulties in acquiring Theory of Mind (Apperly et al., 2010; Baron-Cohen et al., 1985; Wellman, 1992, a.o.) to back up her observation of the preference for *de re* over *de dicto* interpretations. That is, the observation that children tend to focus on reality and the broader context and find it difficult to reason about others' minds helps explain why (even) adults have a preference for *de re* that emphasizes what is going on from the actual world and from the global context, compared with *de dicto* which emphasizes mental status. This study in law cites different sources from the Default Semantics framework, but arrives at similar conclusions on the empirical accessibility of the *de re/de dicto* ambiguity.

Finally, evidence for a dispreference for *de re* comes from experimental findings reported in Zhang and Davidson (2021). They designed an acceptability task, exemplified in (6). The evaluating context featured a protagonist Julie who falsely associated the authorship of the target poem with Nicole, but in reality, and from the speaker's perspective, the poem was written

[3] 18 U.S.C. §1503 (2010).
[4] 18 U.S.C. §1512 (b) (2010).

by Elizabeth. According to a similar scenario in Romoli and Sudo (2009),[5] the belief report with the noun phrase *Nicole's poem* should be interpreted *de dicto* and the one with *Elizabeth's poem* should be *de re*.

(6) *Evaluating Context*: Julie is one of the judges of an ongoing poetry competition. The best poem that she has read so far is an extremely intriguing poem about the ocean. She believes that this poem will win the competition. Julie remembers being told that Nicole, one of the best-known poets, submitted a poem about the ocean to the competition. Therefore, Julie concludes that this poem must be written by Nicole, and the first prize will be going to her. However, this poem was actually written by Elizabeth, a younger and lesser-known poet. It is just a coincidence that the two poets wrote about the same topic.
   a. Julie believes that Nicole's poem will win the competition. (*de dicto*)
   b. Julie believes that Elizabeth's poem will win the competition. (*de re*)

In the experiment of Zhang and Davidson (2021), speakers of English read four stories similar to (6). Two stories had a follow-up test sentence with a *de re* reading, and the other two stories had a sentence with a *de dicto* reading. Participants used a continuous slider bar to give their judgment. Transforming the judgment data into the sides of agreement and disagreement, the results show that around 20% of participants disagreed with both *de re* sentences, about 40% disagreed with one of the *de re* sentences, and the remaining 40% agreed with both *de re* sentences. In contrast, almost all of the participants showed strong agreement with the *de dicto* sentences. What is more noteworthy is that a sizable proportion of the disagreement concerning *de re* aggregated on the very edge of the slider bar. In contrast, the majority of the *de dicto* judgments aggregated on the edge of the agreement side. This empirical finding, especially the peculiar bimodal distribution of *de re* acceptability, is surprising in light of other work, such as Romoli and Sudo (2009), where both readings are predicted to be generally acceptable.

Thus, we see several different sources in the literature that seem to report greater ease or difficulty, respectively, in accessing *de re* readings in comparison to *de dicto* readings of noun phrases. This seems like an area ripe for more careful experimental consideration, and, especially, further investigation of the role that context and pragmatics may play in these judgments.

---

[5] The example in Romoli and Sudo (2009) is shown in (i). The context explicitly shows that the belief holder does not think the *de re* noun phrase refers to the target individual, because the belief holder registers the *de dicto* term as associated with the target individual.

(i) John thinks that the president of the United States is smart.
Evaluating context: Consider the situation as of today [2009], in which Barack Obama is the president of the United States, and suppose that John wrongly thinks that Al Gore is. In this context, the sentence has two interpretations. It can be read as reporting John's belief about Barack Obama or about Al Gore. The former is called the *de re* reading and the latter the *de dicto* reading.

1.2 More motivations for experimental investigation

Given the reported disparities in the existing literature between *de re* and *de dicto* interpretations, more carefully controlled experimental research seems prudent; in this section we further motivate an experimental take on this topic.

For one thing, from the perspective of research methodology, the experimental investigation of *de re/de dicto* provides an instance that enriches the discussion on "the nature of empirical evidence in research on meaning" (Tonhauser & Matthewson, 2015, p. 1). To yield stable replicable and transparent data for theoretical development, Tonhauser and Matthewson (2015) argue that one needs to provide not only the linguistic expression under investigation, but also the context in which the expression is uttered, a response by a native speaker to a task involving the linguistic expression in the context, and information about the native speakers that provide the response. Gibson and Fedorenko (2010) also argue for using multiple items with controlled experiments to eliminate confounding factors like specific lexical properties in a single sentence and idiosyncrasies from the contexts (aspects of experimentation that are valuable even in very small scale studies, as discussed in Davidson (2020)). Existing research in semantics and pragmatics has already uncovered valuable insights by studying the influence of contexts, experimental paradigm, and/or response options on linguistic judgments, and has illustrated some of the advantages of becoming more "experimental" (see, e.g. Schwarz et al. (2007) and Jasbi et al. (2019) for scalar implicature and the influence of contexts as well as response options; see Jasbi, Bermudez, and Davidson (2023) and Jasbi, Bermudez, Zhang, et al. (2023) for logical connectives and the effect of experimental paradigm on cross-linguistic findings). Our research is another case study that advocates for rigorous experimental practice in the study of meaning.

Second, empirically investigating the *de re/de dicto* reading acceptability provides a case study of how formal theories of meaning that involve logic and mathematical techniques align with psycholinguistic findings that delineate the psychological representation of meaning. For example, it would be interesting to explore whether there is an analogy between (i) the semantic representation of the *de re/de dicto* ambiguity with the formal scope technique, compared to its empirical acceptability, and (ii) the formal quantificational scope and the related psycholinguistic processing mechanisms in classic scope interactions (see Brasoveanu & Dotlačil, 2019, for a review). We know that the inverse scope reading is sometimes hard to obtain, as in (7) (e.g. Anderson, 2004; Tunstall, 1998). What is more interesting is whether a wider existential scope in *de re* that is opposite to its surface position also leads to difficulty in interpretation or processing.

(7) a. A caregiver comforted Mary every night. (The inverse scope reading is, at every night, there was a different caregiver who comforted Mary.)
b. A boy climbed every tree. (The inverse scope reading is, for every tree, there is a different boy who climbed it.)

Last but not least, the *de re/de dicto* ambiguity in belief reports is one of many phenomena that can be used to investigate the cognitive relation between people and propositions (see Nelson, 2023, for a review). It tries to answer the question of how one's belief is encoded in language, and in turn, how the ambiguity of language leads to varied and nuanced interpretations of one's belief content. It has attracted interdisciplinary discussion from philosophy (e.g. Cohen et al., 2021; Epstein et al., 2023; Lederman, 2022; Richard, 1990), cognitive science (e.g. Apperly & Robinson, 2003; Robinson & Apperly, 2001), and even artificial intelligence (e.g. Wiebe & Rapaport, 1986; Wu et al., 2023). Understanding what contexts facilitate or impede a certain reading of the belief report could shed light on the research agenda that aims to understand the relation between language and mind. More practically, this line of research could also complement existing research about false belief tasks and Theory of Mind (e.g. Baron-Cohen et al., 1985; Wimmer & Perner, 1983) where the focus has been on the predicate of the belief content (e.g. *Sally believes that the marble is in the box.*) rather than the referential properties of noun phrases (e.g. *the marble*) inside belief reports.

1.3 A finer-grained categorization of *de re* contexts and the *de re/de dicto* term

In this section, we set up a finer-grained categorization of (i) the *de re* permitting contexts and (ii) the *de re/de dicto* terminology to better operationalize our experimental investigation. The aim is to collect accurate linguistic judgments for a specific linguistic expression with a clarified meaning in a well-controlled context.

First, we subdivide the *de re* permitting contexts into two types: *the ignorance context* and *the misapprehension context*, adopting the terminology introduced by Sudo (2014). The misapprehension context in Sudo (2014) is shown in (8) where the attitude holder is wrong about the identity of the referent (Sudo does not give an ignorance context example).

(8) John thought that the linguist was nervous.
Evaluating context: John interviewed two girls, Mary and Sue. He was informed beforehand that one of them is a linguist, but was not told which. We know that Sue is the linguist. After the interviews, John wrongly concluded that Mary was the linguist, because she said she speaks five languages. John thought that Sue was nervous.

In general, the misapprehension context applies to cases where the belief holder is wrong about the identity of a referent and assigns to it a noun phrase that is true in the belief holder's mind but false in reality. The *de re* context in the poetry competition example in (6) features such a misapprehension context. Judge Julie believes the poem was written by Nicole and assigns the noun phrase *Nicole's poem* to it. But in reality and the broader story context, the poem was written by Elizabeth (the latter noun phrase *Elizabeth's poem* is read *de re*). On the other hand, the ignorance context applies to cases where the belief holder is ignorant of the association between the *de re* noun phrase and the referred object. The context in the abstract reviewing example in

(6) that permits the *de re* reading can be categorized as the ignorance context, since the belief holder John is not aware that the abstract was written by the addressee.

Consistent with most formal semantics literature on the *de re/de dicto* ambiguity, Sudo (2014) does not predict or discuss differences between these two contexts in the acceptability of the *de re* interpretation of definite noun phrases. Yet in Zhang and Davidson (2021), the judgment data shows that the misapprehension context prioritizes *de dicto* and disfavors *de re*. One potential hypothesis is that the misapprehension context highlights the contrast between the belief holder's mental state and the story's broader context. When the target sentence *Julie believes that Nicole's poem*$_{de\ dicto}$ *will win the competition* starts with *Julie believes…*, it highlights the belief holder's mental state. Since the *de dicto* term is consistent with the content of the belief holder's mind and, thus, is consistent with the preamble, it is prioritized over the *de re* noun phrase. Previous research on children's interpretation of belief reports shows that highlighting the belief state of the protagonist in the context increases the likelihood that children attend to the belief content (Lewis et al., 2017). It is thus interesting to see whether the same mechanism is playing a role in the disproportionate judgment pattern in *de re/de dicto* noun phrases.

Secondly, we adopt a three-way taxonomy of the *de re/de dicto* ambiguity, as referential *de re*, referential *de dicto*, and attributive *de dicto*. This is motivated not only by the fact that ontologically two types of contexts emerge which admit the *de dicto* reading of definite noun phrases, but also that, from an experimental perspective, we could have better control over what context corresponds to what specific readings of the sentence. Based on previous discussion of the referential and attributive properties of definite noun phrases (Donnellan, 1966; Fodor, 1970; Jaszczolt, 1997), we introduce this tripartite taxonomy using the poetry competition example, repeated here as (9).[6]

(9) Julie believes that Elizabeth's poem will win the competition.

The context that assigns Elizabeth's poem a <u>referential de re</u> reading: Julie does have a particular poem in mind that she believes will win the competition. However, Julie doesn't recognize the description *Elizabeth's poem* as a description of the poem she has in mind.

The context that assigns Elizabeth's poem a <u>referential de dicto</u> reading: Julie does have a particular poem in mind that she believes will win the competition. She has the poem in mind as *Elizabeth's poem*. However, in reality, the poem is Nicole's poem. In other words, Julie falsely represents the poem by using a referential term that is false in the actual world.

[6] Note that there is a fourth reading here: Julie believes that poem A will win the competition and knows that poem A was written by Elizabeth. This reading is not controversial, and less interesting from a theoretical sense, so we didn't put it in parallel with the other three. More nuanced readings are pointed out in the literature, such as *Elizabeth* is *de re* but *poem* is *de dicto* (Charlow & Sharvit, 2014) or the terms "third" and "fourth" reading in Fodor (1970). We do not touch upon those readings for now.

> *The context that assigns Elizabeth's poem an attributive de dicto reading*: Julie does not have any particular poem in mind, but simply believes that whichever poem written by Elizabeth will win the competition.

We can achieve a precise and finer-grained understanding of how the distinction of *de re/de dicto* in definite noun phrases can be mapped to specific readings in specific contexts by integrating the two context types and the three readings of intensional definite noun phrases. The ignorance context only permits a referential *de re* reading of the corresponding definite noun phrase. The misapprehension context (theoretically) permits a referential *de re* reading of one definite noun phrase and a referential *de dicto* reading of another (competing) definite noun phrase. The following experiments aim to test whether the division of context types and terminologies can help address the bimodal distribution of acceptability judgments of *de re* reported in prior work by Zhang and Davidson (2021).

1.4 Experiment outline

This section introduces the outline of three experiments. All three experiments adopted an acceptability task with adult native speakers of English.

Experiment 1 replicated Zhang and Davidson (2021), and found that when the context featured misapprehension and allowed a referential *de re* reading for one noun phrase and a referential *de dicto* reading for another, the referential *de re* reading showed a bimodal distribution of acceptability judgments.

Experiment 2 put the critical sentence from Experiment 1 under contexts that supported only an attributive *de dicto* reading (the control context) and contexts that supported only a referential *de re* reading (the ignorance context). There, we found no statistical difference in acceptability ratings between the two readings – both received high agreement. This shows that the referential *de re* interpretations of belief reports are acceptable in the ignorance contexts.

Experiment 3 tested the hypothesis that the misapprehension context, i.e. the co-existence of a referential *de re* definite noun phrase and a referential *de dicto* definite noun phrase in the same context, would make the former reading less acceptable. By juxtaposing the misapprehension context from Experiment 1 and the ignorance context from Experiment 2, we found evidence supporting this hypothesis.

In sum, we show that while the *de re* reading of definite noun phrases is undoubtedly allowed by the grammar, its acceptability is vulnerable to contextual factors relating to competition between multiple possible referential expressions. When a competing referential *de dicto* noun phrase appears in the same context, in other words, when the belief holder associates a wrong term with the target object, the *de re* reading with a different term becomes much less acceptable. We discuss the potential mechanisms and implications in Section 5.

2. Experiment 1

Experiment 1 aims to replicate Zhang and Davidson (2021), asking whether the bimodal distribution of *de re* judgments found in that study would persist in another round of testing with slight modifications of the response type that will be used in the rest of the studies in this article.

2.1 Methods

2.1.1 Participants

60 participants who self-identified as English monolinguals from the United States (aged 39.15 ± 12.05) took this study. They were recruited from the online crowdsourcing platform Amazon's Mechanical Turk and were paid $2.00 for their participation.

2.1.2 Materials and procedures

Since Experiment 1 was a replication of Zhang and Davidson (2021), the experimental materials were the same. The only difference was that we adopted a discrete fully-labeled Likert scale to collect the judgments, rather than a continuous slide bar, as in the original study. We chose a Likert scale over a binary option or a continuous slider based on the following considerations: First, Likert scales provide finer-grained levels that can reveal potential judgments that would otherwise remain concealed on a binary scale (e.g. Jasbi et al., 2019; Katsos & Bishop, 2011; Zhang et al., 2021), and can achieve the same statistical power with a smaller sample size (Cremers et al., 2023); second, the labels at the intermediate levels (e.g. "somewhat agree", "uncertain", "somewhat disagree") potentially offer better interpretability than a continuous slider bar when it comes to mapping participants' intermediate choices with their actual interpretations; third, choosing a Likert scale does not result in a significant loss of sensitivity compared to the continuous slider bar (Marty et al., 2020; Sprouse & Almeida, 2017).

As for the specific experimental design, participants read four stories (113 ± 6.4 words) in a Qualtrics survey and gave their acceptability judgments on a subsequent target declarative sentence in terms of how accurately each sentence reflected the facts in each story. The story and the sentences appeared on separate pages. There was no time pressure to complete the experiment. Participants could return to any story and change their answers at any time before submitting their answers.

Each story portrayed a protagonist who holds a belief toward a person or object (we simplify this using *entity*). The story fosters a misapprehension context where the protagonist falsely takes one definite noun phrase to refer to the target entity but in reality, the correct attribution should come from the other definite noun phrase. One of the trials is shown in **Table 1**,[7] which

[7] The critical sentences that determined the condition are italicized in the table for illustrative purposes. They were not italicized in the actual experiment. All experiment stimuli can be accessed through the OSF link https://osf.io/6pvdz/.

is the same as (6). Julie falsely believes the poem was written by Nicole. In reality, it was written by Elizabeth, which Julie is unaware of. Given this context, the target sentence to be judged featured a report describing the protagonist's belief. The experimental manipulation concerned what definite noun phrase was used in the belief report as the referring expression. Using the definite noun phrase held to be true in the protagonist's mind would render the belief report a referential *de dicto* reading. In contrast, using the definite noun phrase held to be true in the broader story context and from the speaker's perspective would render the report a referential *de re* reading. In theory (Romoli & Sudo, 2009; Sudo, 2014), both readings are predicted to be true.

Additionally, for each story, there were three sentences accompanying the target sentence as fillers and controls; one of these was true given the context, one was false, and the third one was undecided because of the lack of verifying information. For each sentence to rate, participants were asked to map their judgment onto a five-point fully labeled Likert scale. Participants gave their judgment depending on whether and to what degree they agreed with the content of the sentence given the story context. By comparing the proportion of different levels of agreement between conditions, we expected to approximate the representative judgment distribution of the two belief report versions.

Table 1: Example story from Experiment 1.

Context Julie is one of the judges of an ongoing poetry competition. The best poem that she has read so far is an extremely intriguing poem about the ocean. She believes that this poem will win the competition. Julie remembers being told that Nicole, one of the best-known poets, submitted a poem about the ocean to the competition. Therefore, *Julie concludes that this poem must be written by Nicole and the first prize will be going to her. However, this poem was actually written by Elizabeth, a younger and lesser-known poet.* It is just a coincidence that the two poets wrote about the same topic.
Instruction According to this story, please indicate to what extent you agree or disagree with the following statements.
Target sentence – Condition 1 Julie believes that **Nicole's poem** will win the competition. **(Referential *de dicto*)** **Target sentence – Condition 2** Julie believes that **Elizabeth's poem** will win the competition. **(Referential *de re*)** Highly Disagree ○ Somewhat Disagree ○ Uncertain ○ Somewhat Agree ○ Highly Agree ○

Altogether, each participant read four stories. We chose four items, because this number was chosen in Zhang and Davidson (2021), and it was easier to compare item differences with a smaller set of scenarios.

Furthermore, the condition manipulation was within-subjects, and participants read stories in both conditions. Across the four stories, two were randomly assigned to be referential *de dicto* and the other two were referential *de re*. The order of stories and the sentences within a story were randomized. To achieve a Latin Square design, we manually created six lists[8] and each participant was randomly assigned to one of the lists. During the actual experiment, participants started by completing three practice trials (sentences to be judged without contexts) to familiarize themselves with the experimental design. In the end, participants completed a survey and provided their demographic information and their self-reported linguistic profile (e.g. reading and writing proficiency, knowledge of other dialects/languages).

2.2 Results

We only analyzed the judgments from participants who passed the practice trials and whose judgments of the fillers were correct more than 75% of the time, so 51 out of 60 participants (85%) contributed their data to our final analysis.[9]

Figure 1 shows that in the referential *de dicto* condition, the majority of judgments goes to the "highly agree" side. This contrasts with the referential *de re* condition, where around 25% of the judgments are "highly disagree", with the majority of choices still going to "highly agree".

We fit the judgment data into Bayesian multilevel cumulative ordinal models, using the *brms* package (Bürkner, 2017, 2018) in R. The five-point judgments were the dependent variable, with non-equidistant intervals between levels on the Likert scale. They were coded from 1 to 5, where 1 indicated "highly disagree" and 5 indicated "highly agree". The two critical condition levels were entered as a dummy-coded fixed effect (reference level = referential *de dicto*). The story was also entered as a fixed effect, and the interaction between the two fixed effects was also included.[10] Random intercepts and slopes for the full fixed effects for the subject were included as random effects to obtain the maximal random effects required for mixed-effects models (Barr et al., 2013). The prior distributions for all the intercepts and coefficients of fixed effects were fitted

[8] We created six lists, because there are six combinations where two stories are randomly interpreted *de re* and the other two are *de dicto*. That is {AB/CD, AC/BD, AD/BC, BC/AD, BD/AC, CD/AB} where the first two stories corresponded to a referential *de dicto* condition and the latter two corresponded to a referential *de re* condition.

[9] The reason why we included the practice trials in the participant screening procedure was that we explicitly asked the participants to choose, e.g. "highly agree", over other choices. A failure to do so indicated a lack of attention to our materials and instructions.

[10] A motivation was that we could derive more insights into story-specific effects on the acceptability of *de re* readings from the statistical analysis.

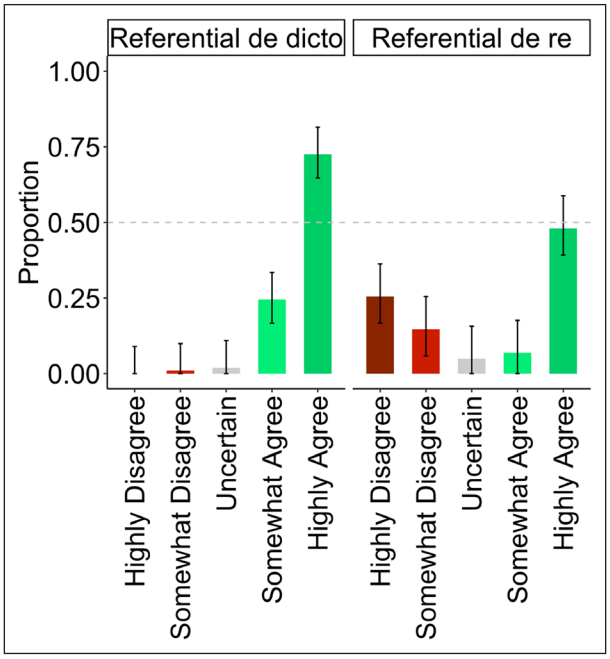

Figure 1: Proportion of different Likert scale choices in the referential *de dicto* and the referential *de re* conditions (the 95% CIs were output from the *MultinomialCI* package (Sison & Glaz, 1995)).

to a normal distribution with the mean as 0 and the standard deviation as 2 (i.e. *Normal* (0,2)*;* the prior for the correlation matrices was set to be *LKJ*(2).[11] The variances for the correlation matrices were set as the default in R. The priors mildly restricted the possible coefficient for each parameter, but still allowed reasonably large variance. The model had four sampling chains, each with 4000 iterations. The first 2000 samples were taken as a warmup. An $\hat{R}$ close to 1.0 marks the convergence of the sampling chain to the underlying posterior distribution of the target predictor (Gelman & Rubin, 1992). The parameter setup also followed previous acceptability rating tasks in psycholinguistics (e.g. Paape et al., 2020; Zhang et al., 2023).

All $\hat{R}$s for the sampling chains for all fixed effects were 1.0, indicating successful convergence. We used the package *emmeans* (Lenth et al., 2018) to evaluate the main effect exerted by the *de re/de dicto* manipulation and the judgment distinction in each story setting. Here we use β to refer to the coefficient estimate and HPD, i.e. highest posterior density, to refer to the shortest interval with the highest density in the posterior distribution of the target coefficient (Box & Tiao, 2011).

Overall, the referential *de dicto* condition receives more agreement than the referential *de re* ($\beta = 1.66$, HPD = [0.16, 3.12]). Interestingly, the agreement distribution of the two conditions

[11] LKJ has been the default weakly informative prior for correlation matrices in *brms*, following Lewandowski et al. (2009) and Nalborczyk et al. (2019).

varied by the story: in story A, there does not seem to be a difference (β = 1.40, HPD = [–0.16, 2.99]); there also does not seem to be a difference in story B (β = –1.30, HPD = [–3.84, 0.95]); in both story C and story D, the referential de dicto reading has a higher agreement rating than the referential *de re* reading (C: β = 3.48, HPD = [1.33, 5.98]; D: β = 3.10, HPD = [0.65, 5.68]). The by-item distribution can also be seen in **Figure 2**.

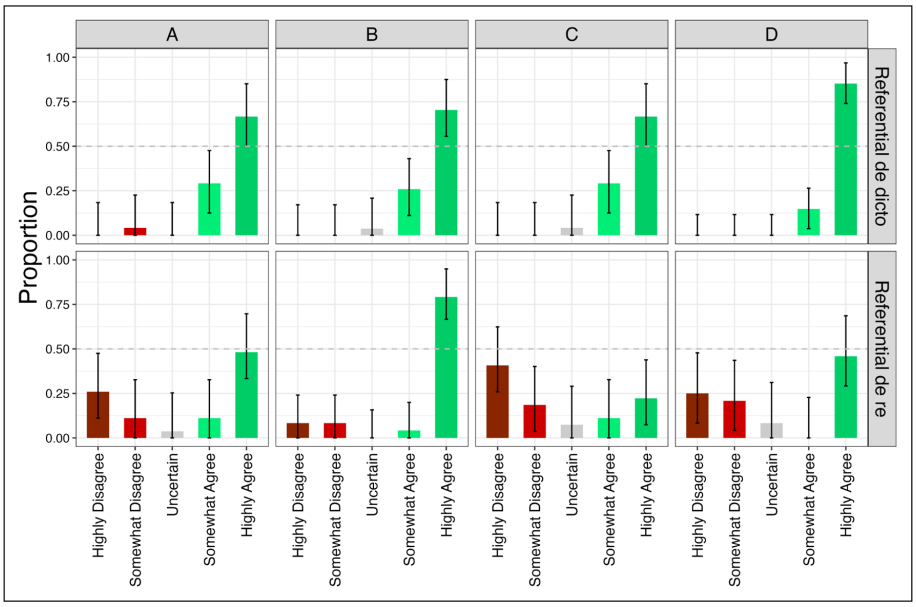

Figure 2: Judgment distribution across the four stories in Experiment 1.

When it comes to individual differences, we see that for the referential *de dicto* condition, more than 50% of the participants chose “highly agree” for both trials; nobody chose “highly disagree” and only one participant chose “somewhat disagree” once. In comparison, only 15 out of 51 participants (29.4%) chose “highly agree” twice for the referential *de re* condition; 5 participants (9.8%) chose “highly disagree” twice. It is clear that participants showed more disagreement and chose more intermediate options for the referential *de re* readings.

2.3 Discussion

In this experiment, we made use of contexts that should support, through two possible means of description, both the referential *de dicto* interpretation and the referential *de re* interpretation of belief reports and varied the target sentence to gauge the judgment distribution of the two readings. We found that while the referential *de dicto* reading was overwhelmingly agreed, the referential *de re* reading led to a bimodal pattern of judgments. This successfully replicated Zhang and Davidson (2021), but uncovered judgment patterns that are new with respect to other literature on this topic, such as Sudo (2014).

The by-item investigation shows that while all four stories received almost identical rating patterns for the referential *de dicto* reading, they exhibited different patterns for the referential *de re* reading. Specifically, in story B, the proportions for "highly disagree" and "somewhat disagree" were the lowest by comparison, replicating the findings in Zhang and Davidson (2021). We speculate that the unique judgment pattern for the *de re* sentence in story B, i.e. *Mrs. Jackson believes that Grace's gift was sent by Mike*, might be related to the information structure of the passive complement clause (see Appendix Table 1 for the complete experimental details). Since passives could (i) highlight the relative newness of the information in the *by* phrase – it was Mike, not someone else, who sent the gift, and/or (ii) emphasize that the subject is affected by the action denoted by the verb – the gift was sent, not received (Ambridge et al., 2016; Pullum, 2014), this passive construction could potentially modulate the information flow so that readers might focus on verifying information in the predicate and ignore the subject in the belief report. Nevertheless, story D also featured a passive structure in the complement clause, i.e. *Tracy believes that Alice's spare apron needs to be washed*, but there is still a significant proportion of disagreement on *de re*. We speculate that the *by* phrase in story B could play a role here, and we leave to future work the further investigation of the effect of passives as well as the information structure of the complement clause on the judgments of the referential *de re* reading.

The investigation of individual differences shows that while more than half of the participants had no problem accessing the referential *de dicto* condition, only 30% accessed the referential *de re* condition plus 10% going for the opposite truth-value judgment.

In sum, the bimodal distribution of the referential *de re* reading was replicated and observed to be systematic. Future studies are needed to disentangle the effects of linguistic information structure and individual differences on *de re* judgments.

3. Experiment 2

It is clear that the contexts theoretically permitting the referential *de re* reading in Experiment 1 featured misapprehension of the belief holder, and in this sense differed from the common *de re* contexts that usually feature an ignorance context (e.g. the *Sleeping Beauty* case in (1) and the conference abstract case in (3)). Besides, to the best of our knowledge, there has been no experimental research that lays out the judgments of the canonical *de re/de dicto* paradigm in a systematic way. Therefore, to collect judgments under a simple canonical *de re* permitting context, with the canonical *de dicto* permitting context as a comparison, Experiment 2 juxtaposed a context that only permits a referential *de re* reading of a definite noun phrase with another context that only permits the attributive *de dicto* reading.[12]

[12] The reason we tested the attributive *de dicto* instead of the referential *de dicto* reading of definite noun phrases in Experiment 2 was that the attributive reading resembles the *de dicto* reading of an indefinite noun phrase, which is closer to the canonical interpretation of *de dicto* (please compare (1) and (9)).

3.1 Methods

3.1.1 Participants

66 participants took this study who self-identified as English monolinguals from the United States (aged 32.94 ± 10.25). They were recruited from the online crowdsourcing platform Prolific and were paid $2.00 ($12–15/hr) for their participation.

3.1.2 Materials and procedures

The experimental design and materials were very similar to Experiment 1, except that the manipulation took place in the context, not at the sentence level, and each context theoretically permitted only one reading. Additionally, the readings to be tested were attributive *de dicto* and referential *de re*. Each story (81.14 ± 13.75 words) portrayed a protagonist who held a belief. The target sentence to be judged featured a belief report. We created two conditions, the attributive *de dicto* condition and the referential *de re* condition, by varying the context of the story in which the same target sentence was to be evaluated. **Table 2** exhibits an example and the full list of the materials is in Appendix Table 2. In the attributive *de dicto* condition, the protagonist Julie believes that whichever poem written by Elizabeth will win – the noun phrase *Elizabeth's poem* does not refer to any specific individual entity in the mind of the speaker, only the contents of Julie's mind. In the referential *de re* context, Julie believes of a particular poem (which exists, according to the speaker) that it will win the competition, but does not know that the author of this poem is Elizabeth. The *de re* context in Experiment 2 was the ignorance context. Additionally, for each story, there were three sentences accompanying the target sentence as fillers and controls; one of these was true, given the context, one was false, and the third one was undecided, because of the lack of verifying information.

For each sentence they were asked to rate, participants were asked to map their judgments onto a five-point fully labeled Likert scale. Crucially, the manipulation of conditions was within-subjects, and participants read stories in both conditions. Across the four stories, two were randomly assigned to be attributive *de dicto* and the other two were referential *de re*. The randomization, the counterbalance treatment, and the experimental procedure were the same as in Experiment 1.

3.2 Results

We only analyzed the judgments from participants who passed the practice trials and whose judgments of the fillers were correct more than 75% of the time. As a result, 60 (90.9%) participants contributed to the analysis.

Figure 3 shows that in the attributive *de dicto* condition, more than 75% of the judgments fall within the "highly agree" category and the distribution is strongly skewed towards the agreement edge. In the referential *de re* condition, more than half of the judgments aggregate

Table 2: Example story from Experiment 2.

Condition 1: Attributive *de dicto*
Julie is a judge of an ongoing poetry competition. She is told that Elizabeth Johnson, one of the best-known poets in the US, submitted a poem to the competition. Julie is a huge fan of Elizabeth. Even though Julie is blind to the authors and does not know which poem is written by Elizabeth, *she believes that no matter which poem Elizabeth submitted, it will win the competition*.
Condition 2: Referential *de re*
Julie is a judge of an ongoing poetry competition. *She encounters an extremely well-written poem and believes that this poem will be the winner of the competition*. This poem happens to be written by Elizabeth Johnson, a well-known poet in the US. But unfortunately, as a judge, *Julie is blind to the authors and therefore does not know it is Elizabeth Johnson who wrote this excellent poem*.
Instruction
According to this story, please indicate to what extent you agree or disagree with the following statements.
Target sentence
Julie believes that Elizabeth's poem will win the competition.
Highly Disagree ○ Somewhat Disagree ○ Uncertain ○ Somewhat Agree ○ Highly Agree ○

to the "highly agree" category, with a similar skewness pattern. By visual comparison, the proportion of judgments from "highly disagree" to "somewhat agree" in the *de re* condition is slightly larger, indicating that the *de re* interpretation might be less acceptable, although, in general, both readings are overwhelmingly acceptable.

We fit the judgment data into Bayesian multilevel cumulative ordinal models. The condition, the story, and their interaction were entered as fixed effects; random intercepts and random slopes for the full fixed effects structure for the subjects were entered as random effects. The prior setting, as well as all the other parameters, were the same as in Experiment 1.

The Bayesian model shows that all $\hat{R}$s for the sampling chains for all fixed effects were 1.0, indicating successful convergence. There was no difference between the attributive *de dicto* condition and the referential *de re* in their agreement distribution ($\beta = 0.825$, HPD = [–0.18, 1.76]) and only in story A was there a difference in judgments of the two conditions ($\beta = 1.31$, HPD = [0.069, 2.63]).

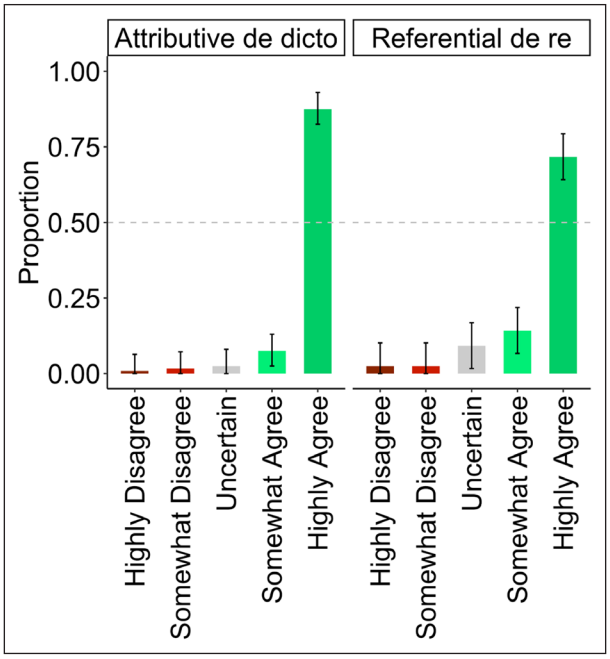

Figure 3: Proportion of different Likert scale choices in the attributive *de dicto* and the referential *de re* conditions. (Error bars indicate 95% multinomial confidence intervals, calculated by the R package *multinomialCI*.)

3.3 Discussion

Experiment 2 found no statistical difference between the referential *de re* reading and the attributive *de dicto* reading of the definite noun phrase in belief reports. This is consistent with prior literature and the general agreement in the field about the *de re/de dicto* ambiguity of noun phrases in belief reports. The comparison between Experiments 1 and 2 reflects the fact that the unacceptability of the *de re* reading of definite noun phrases is conditional on the context. More specifically, we know that the *de re* context in Experiment 1 featured the misapprehension of the belief holder, while the *de re* context in Experiment 2 featured the ignorance of the belief holder with respect to the *de re* expression.

4. Experiment 3

Experiment 3 was designed in order to test more directly whether the *de re* reading of definite noun phrases was degraded in the misapprehension context, compared to the ignorance context. To do this, we juxtaposed contexts that only supported the referential *de re* reading, as in Experiment 2 (the ignorance context), and contexts that should theoretically support both the referential *de dicto* and *de re* readings, as in Experiment 1 (the misapprehension context), in a within-subjects design. In both conditions, participants were asked to just rate the sentence

with the *de re* interpretation. If this specific context setup affects the acceptability of *de re*, we would expect that in contexts that mirrored Experiment 1, the bimodal distribution would still persist; in contexts that mirrored the *de re* condition in Experiment 2, there would be no to few disagreements for the *de re* readings.

4.1 Methods

4.1.1 Participants

60 participants who self-identified as native speakers of English from the United States (aged 33.02 ± 8.35) were recruited from Prolific. They were paid $2.00 for their participation.

4.1.2 Materials and procedures

Table 3 shows an example story. In the referential *de re* condition (the ignorance context), there is only one valid nominal expression referring to the target object, but the belief holder is unaware of such a relation; the scenario only supports a referential *de re* interpretation of the belief report. In the referential (*de dicto* + *de re*) condition (the misapprehension context), there are two valid nominal expressions: one has a *de re* interpretation, because the protagonist is unaware of the association between the expression and the target object; the other has a referential *de dicto* reading which the protagonist associates with the object in her mind but is wrong in the broader story context. Please see the full list of materials in Appendix Table 3.

Table 3: Example story from Experiment 3.

Condition 1: Referential *de re* Julie is a judge in an ongoing poetry competition. She encounters an extremely well-written poem and believes that this poem will be the winner of the competition. This poem happens to be written by Elizabeth Johnson, a well-known poet in the US. But unfortunately, as a judge, *Julie is blind to the authors and therefore does not know it is Elizabeth Johnson who wrote this excellent poem.*
Condition 2: Referential (*de dicto* + *de re*) Julie is a judge in an ongoing poetry competition. She encounters an extremely well-written poem about the ocean and believes that this poem will be the winner of the competition. Julie remembers being told that Nicole, one of the best-known poets, submitted a poem about the ocean to the competition. *Therefore, Julie concludes that this poem must be written by Nicole and the first prize will be going to her. However, this poem was actually written by Elizabeth, a younger and lesser-known poet.* It is just a coincidence that the two poets wrote about the same topic.

(Contd.)

Instruction
According to this story, please indicate to what extent you agree or disagree with the following statements.

Target sentence
Julie believes that **Elizabeth's poem** will win the competition. (**Referential *de re***)

Highly Disagree	Somewhat Disagree	Uncertain	Somewhat Agree	Highly Agree
○	○	○	○	○

With this context manipulation, participants read the story and judged the following belief report, where the nominal expression inside the belief report was *de re*. There were four scenarios. The counterbalance and the randomization design were kept the same as in Experiments 1 and 2.

4.2 Results

54 out of 60 participants (90%) contributed to the crucial analysis, after the same screening procedure. Aligning with the prediction, **Figure 4** shows that the bimodal distribution only appeared in the referential (*de dicto* + *de re*) condition; in the referential *de re* condition, the majority of the judgments accrued on the edge of agreement. This finding was also supported by Bayesian multilevel cumulative ordinal models. Here, both the dummy coded condition (reference level = (*de dicto* + *de re*)) and the story (reference level = a) as well as their interaction were entered as fixed effects; random intercepts and random slopes for the full fixed effects structure for the subject were entered as random effects. The priors and all the meta parameters were set the same as in previous experiments. All $\hat{R}$ for the sampling chains for all fixed effects were 1.0, indicating successful convergence. The result shows that, overall, the *de re* only condition elicited more agreement than the (*de dicto* + *de re*) condition (β = 1.21, HPD = [0.15, 2.71]).

Figure 5 exhibits the by-story judgment pattern between the two conditions. We see clearly that within the context designed to support both the referential *de dicto* and referential *de re* reading, there are larger proportions of disagreement on the target sentence, compared to the *de re* only condition. The statistical analysis shows that in story A and story C, in the (*de dicto* + *de re*) condition, there was marginally more disagreement than in the *de re* only condition (story A: β = 1.06, HPD = [–0.11, 2.54]; story C: β = 1.29, HPD = [–0.19, 3.32]). In story B, there wasn't a significant difference between the two conditions (β = 0.22, HPD = [–1.79, 2.60]). In story D, there was a statistical significance between the *de re* only ratings and the (*de dicto* + *de re*) ratings (β = 2.32, HPD = [0.58, 4.70]). The peculiarity of story B persisted here.

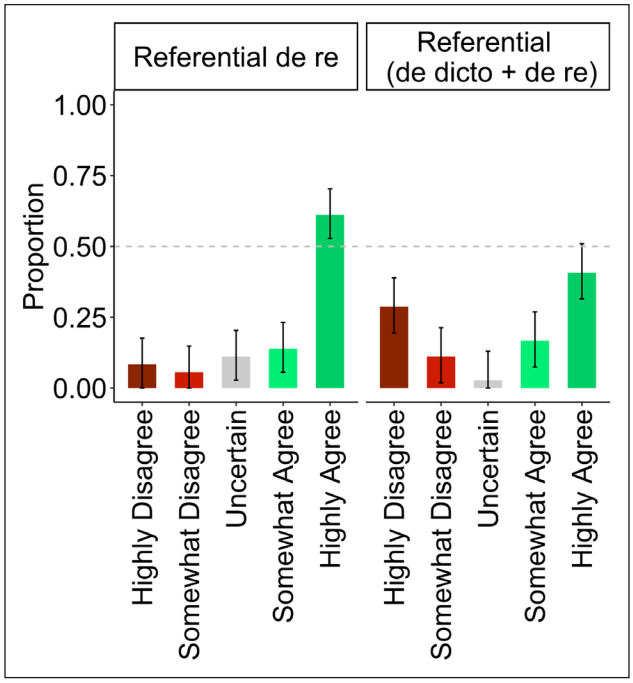

Figure 4: Proportion of different Likert scale choices of the referential *de re* reading in two contexts (the 95% CIs were output from the *MultinomialCI* package). The referential *de re* context refers to the ignorance context. The referential (*de dicto* + *de re*) context refers to the misapprehension context.

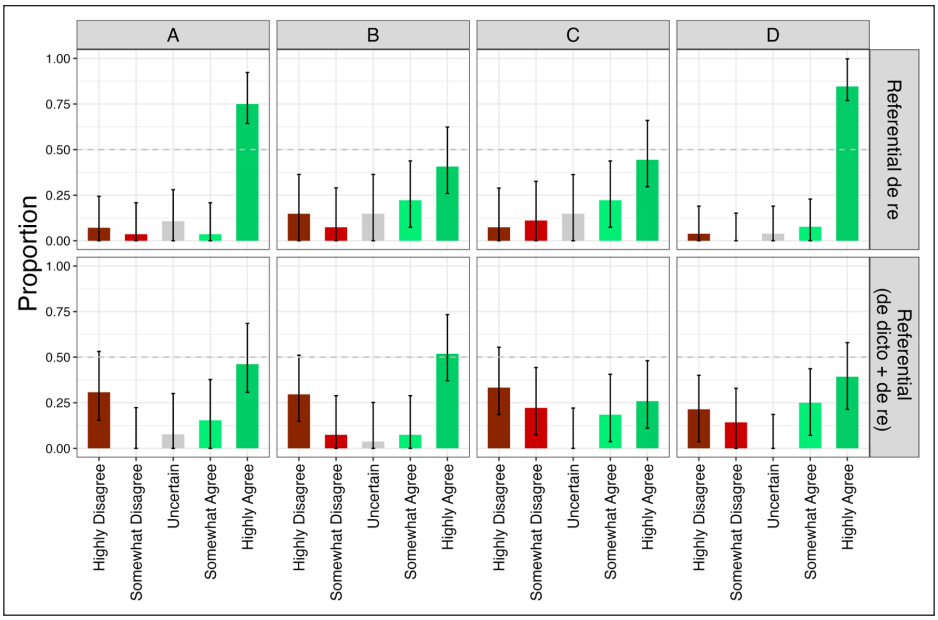

Figure 5: Acceptability ratings on the two conditions across four stories in Experiment 3.

Similar to Experiment 1, we also observe a difference in the percentage of participants choosing edge judgment labels between the two conditions. In the referential *de re* condition, 18 out of 54 participants (33%) chose "highly agree" twice and only 1 participant chose "highly disagree" twice. In the referential (*de re* + *de dicto*) condition, only 14 (26%) stayed with the "highly agree" option twice, but 9 out of 54 participants (16.7%) went to "highly disagree" twice. This confirms that participants' judgment behavior changed between conditions.

4.3 Discussion

Experiment 3 further supports our hypothesis. The *de re* readings of definite noun phrases are acceptable in contexts that uniquely support a *de re* reading with an ignorance scenario. The acceptability of *de re* is affected in contexts which also permit a referential *de dicto* reading of a competing noun phrase with a misapprehension scenario. The item-wise differences persist in this experiment, which further supports the claim that the *de re* bimodal distribution arises in the face of this particular contextual manipulation.

Furthermore, there seem to be more disagreements on *de re* under the ignorance context in Experiment 3 than in Experiment 2 (**Figure 4** vs. **Figure 3**).[13] We speculate that a priming effect (Bock, 1986) could play a role here: while participants chose "disagreement" for the *de re* interpretation in the misapprehension context (i.e. the referential (*de re* + *de dicto*) context in **Figure 4**), they were more tempted to do so with *de re* in the ignorance context (i.e. the referential *de re* context in **Figure 4**). Since priming effects have been found not only in syntactic processing (e.g. Tooley & Traxler, 2010), but also in semantic and pragmatic reasoning (Bott & Chemla, 2016; Raffray & Pickering, 2010; Rees & Bott, 2018), it would be interesting to see whether and how they can be applied to the *de re/de dicto* context in further controlled studies.

5. General discussion

One of the biggest challenges in modeling the meaning of a linguistic expression is to understand what aspects of meaning are stable across contexts and what aspects are context-dependent. In this study, we applied an offline truth-value judgment task to investigate the contextual effect on the acceptability of *de re* readings of definite noun phrases, with the goal of understanding semantic vs. pragmatic factors in *de re* interpretations. In Experiment 1, we replicated the bimodal distribution of the acceptability of *de re* readings in previous research. In Experiment 2, we found that by changing the context into the canonical *de re* permitting context, the *de re* readings did not suffer from degradation. Inspired by Sudo (2014), we labeled the contexts as the ignorance context and the misapprehension context. We found that when the *de re* reading of a definite noun phrase was evaluated under the ignorance context – where the context only

[13] We thank one reviewer for pointing this out.

allowed one noun phrase to refer to the target object and the belief holder did not have that noun phrase in mind – the *de re* reading was almost always accepted; when the context featured a misapprehension scenario – where the belief holder wrongly associated a competing noun phrase with the object – the *de re* reading of the actually correct noun phrase in the belief report would receive a bimodal distribution of judgments. Altogether, these results suggest that the *de re* reading of definite noun phrases is generally available in English, but that its acceptability is vulnerable to contextual effects, particularly the competition with a *de dicto* interpretation in cases of misapprehension.

Why might the misapprehension context elicit judgments in favor of the referential *de dicto* reading of definite noun phrases and against the *de re* reading? We provide our tentative explanation by integrating the concept of pragmatic alternatives with the incremental processing mechanism in psycholinguistics.

We know that alternatives play a critical role both in assessing truth conditional semantics and in tightly related pragmatic processing, where the specific language and the larger pragmatic context constrain the alternatives that are considered. For example, if we consider the use of alternatives for definite descriptions in an anaphoric environment, we see that the language constrains possibilities for reference: some languages allow entirely covert noun phrases, while others do not; some mark definite determiners, while others have covert definite determiners; etc. For instance, to express the meaning "I bought a book$_i$. BOOK$_i$ was expensive.", the specific linguistic construct (e.g. definite noun phrases, pronouns) for BOOK, which can be used anaphorically to refer to the book bought by me, is determined by the available alternative referential nominal expressions in a language and the specific anaphoric constraints in that language's grammar (Ahn, 2020). In Mandarin, the subject "BOOK$_i$" could be covert or a bare noun, which influences the relative prominence of other anaphoric expressions. But the lack of these two alternatives in English would render other anaphoric expressions like the definite noun phrase with an explicit definite article, *the book*, more prominent. Another example of linguistic constraints interacting with contextual factors is the well-studied case of processing scalar implicature. As shown by Degen and Tanenhaus (2016), the computation of scalar implicature (i.e. *some* is inferred as meaning *not all*) is affected by the availability of context-specific alternatives. If numerical quantifiers such as *two* or *three* appeared as alternatives for *some* and *all* to describe the quantity of potential objects, the computation of scalar implicature for *some* would be prolonged.

Following the alternatives hypothesis, the misapprehension context in the case of *de re/de dicto* simultaneously provides two linguistic means of referring, one the *de dicto* noun phrase and the other the *de re* noun phrase, which automatically sets these up as competing alternatives.[14]

[14] The uniqueness of this case is that while the anaphoric referring expressions in Ahn (2020) and the implicature reading of *some* in Degen and Tanenhaus (2016) deal with systematic competition involving function words (e.g.

During sentence comprehension, readers in the misapprehension context need to critically analyze their differences. The referential *de dicto* noun phrase is an appropriate referential expression only in the context of the belief holder's mind, not in the context of the broader story. The referential *de re* noun phrase is appropriate for reference in the broader context of the story, not in the context of the belief holder's mind.

When the participant reads the preamble of the critical statement from left to right, e.g. *Julie believes that …*, in the poetry competition story, an incremental parser might anchor the mental representation of the sentence to the belief holder's mind and build a discourse structure model that expects further discussion relevant to what the belief holder believes to be true. The upcoming *Elizabeth's poem*, with a referential *de re* reading which is evaluated to be true only in the broader context, not in the belief holder's mind, would require a revision in the reader's mental model, and thus lead to a degraded linguistic judgment. This might be analogous to discussions of the difficulty of getting the inverse scope reading for classic quantification scope sentences like *A boy climbed every tree* (Altmann & Steedman, 1988; Brasoveanu & Dotlačil, 2019; Fodor, 1982): when hearing the preamble *a boy climbed…*, listeners add a boy to the discourse model and that boy stands in the climbing relation with whatever would come up as the direct object. There, a surface scope reading ("there is a unique boy who climbed every tree") could arguably be more accessible than the inverse scope reading ("for every tree, there is a different boy who climbed it") because the subsequent *every tree* naturally and coherently builds up the original discourse model with one boy. On the other hand, the inverse scope reading requires the revision of the discourse model to multiple boys, which could result in an interpretation difficulty. Going back to the *de re/de dicto* story, the referential *de dicto* term *Nicole's poem*, which is evaluated to be true in Julie's mental world, is a natural follow-up of *Julie believes…*, compared to the referential *de re* term *Elizabeth's poem*. Along these lines, we hypothesize that it might take longer to process the *de re* sentence under the misapprehension context; future online measurements, such as self-paced reading or eye-tracking, could provide supporting evidence. To better understand how the word-by-word incremental processing mechanism plays a role in the interpretation of *de re*, it would be interesting to test variations of belief reports, such as *Elizabeth's poem will win the competition, Julie believes*, and cross-linguistic variations where the complement clause linearly precedes the matrix clause or the matrix verb. We predict that the *de re* reading would receive more disagreement as long as the linguistic constituents in that sentence are evaluated with respect to different contexts and the overall context features misapprehension, regardless of the linear order of these constituents.

---

determiners, demonstratives, quantifiers, numerals), the competition between a *de re* noun phrase versus a *de dicto* noun phrase is motivated purely by the selection of content words in the referential noun phrases.

A related, perhaps even more speculative, perspective on understanding the interpretive process of this phenomenon comes from the false belief tasks in the Theory of Mind (ToM) literature (e.g. Apperly, 2012; Apperly & Robinson, 2003; Onishi & Baillargeon, 2005; Wimmer & Perner, 1983). We might expect ToM to play some role in resolving this ambiguity, and yet it makes essentially the opposite prediction to our findings. ToM generally suggests that accessing others' mental status is harder and costs extra cognitive resources (e.g. Gopnik, 1993; Wimmer & Perner, 1983), while here it is the referential *de dicto* (belief holder oriented) reading that is always accessible, and it is the *de re* (speaker oriented) reading that has bimodal acceptability. It remains for future work to understand the role, if any, for ToM in this kind of task.

Furthermore, our research tackles the bimodal distribution of judgments of *de re* noun phrases in some contexts, but leaves abundant room for exploration of the theory of *default de re* (Jaszczolt, 1997) and related relevant legal observations (Anderson, 2013). One tentative way to connect the three pieces discussed in Section 1.1 is that the broader scenario setting of *de re* interpretation might affect its accessibility. In the current experiments, the setting requires comprehenders to judge the acceptability of a *de re* reading of an ambiguous belief report, given a specified context. On the other hand, in the settings of Jaszczolt (1997) and Anderson (2013), the parties involved need to decide which interpretation of the *de re/de dicto* ambiguities of a sentence is more salient, without an assumed context or even the knowledge of such ambiguity. This contrast is analogous to a contrast between (i) judging whether the sentence *Jack walked past the bank* is acceptable, when describing a picture in which Jack walks past a financial institution (rather than the river bank) and (ii) providing the most salient interpretation of the sentence *Jack walked past the bank*. We show that the accessibility of the *de re* reading is subject to a well-controlled context, but this does not mean that in any scenario, the *de re* reading is discouraged. In fact, during a rapid conversation or in a pressured environment, participants presumably bring a wealth of top-down cues about the most salient interpretation at hand to their interpretation, in order to fulfill the higher-level goal in social interactions. With this, we suggest the adoption of more diverse research methodologies, for instance, a forced choice between the *de re* and *de dicto* interpretations or a paraphrasing task of ambiguous materials, to explore the salience of different ambiguous readings of belief reports in diverse settings.

Finally, this work is not without limitations. First, more work needs to be done to explain the item variance in Experiments 1 and 3, where some stories appear to receive more disagreement on the referential *de re* reading than others. Currently, one hypothesis concerns the information structure of the embedded clause: a passive structure with the *by* phrase could drive the interpretive focus to the predicate or the *by* phrase (Ambridge et al., 2016; Pullum, 2014) and cause the ignorance of the subtlety in the critical embedded subject, although this hypothesis requires a more well-controlled investigation to test properly. Similarly, individual differences between participants in *de re* disagreement are worth exploring. A valuable research question

here is whether there is a natural way to demarcate groups of semantic comprehenders, who might find both *de re* and *de dicto* equally available, versus pragmatic comprehenders, who are more sensitive to contextual factors for interpreting sentences and may prefer one reading. We might also find a fruitful division of individuals who retain openness to ambiguity versus individuals who find it difficult to switch interpretations once one is found. It could also be helpful to see whether other contextual factors (e.g. something other than a competing *de dicto* term which also highlights the belief holder's mental status, as in Lewis et al. (2017)) can also affect the acceptability of *de re* readings. One potential direction is the Question under Discussion (e.g. Roberts, 2012; Ronai & Xiang, 2021): if the context makes the belief holder's mental activities the main topic of discussion, could the *de dicto* reading be even more prioritized than *de re*?

Overall, this study provides the first comprehensive experimental investigation into the acceptability of *de re/de dicto* readings of definite noun phrases and explores the effect of context on linguistic judgment. We hope this piece of work lays out the empirical foundation to study the referential properties of noun phrases in the intensional domain and enriches the set of linguistic phenomena that have increasingly attracted experimental methodological inspection (together with Jasbi, Bermudez, Zhang, et al., 2023; Jasbi et al., 2019; Tonhauser et al., 2018, a.o.). Our findings also extend the processing of scopes from the classic quantificational scope (Anderson, 2004; Brasoveanu & Dotlacil, 2015; Brasoveanu & Dotlačil, 2019; Tunstall, 1998) to the intensional domain. We hope to see more work along the lines that discuss the relation between formal semantic representations of a language and the mental processes of the speaker/listener (Fodor, 1982). Furthermore, this study also sheds light on the interdisciplinary interest of language and mind. Going beyond the developmental trajectory of ToM, which has shown biases toward a speaker-oriented perspective (Anderson, 2013; Jaszczolt, 1997; Lewis et al., 2017; Wang et al., 2020), we show that contextual manipulations can guide readers toward a preference for making reference in terms of others' mental states over one's own.

Data accessibility statement

The experimental stimuli, raw data, graphs, and processing codes can be viewed and downloaded from the OSF platform via https://osf.io/6pvdz/.

Ethics and consent

All the human participants in the three experiments were recruited from online crowdsourcing platforms. This study has been approved by the Committee on the Use of Human Subjects (CUHS) at Harvard University, which serves as the Institutional Review Board (IRB17-0250: Online experimental semantics studies). Every participant in this study gave their consent before participating in the experiments, and their data were anonymized.

Acknowledgements

We gratefully thank three thoughtful and careful reviewers for their generous feedback that significantly improved this paper, as well as Shannon Bryant, Gennaro Chierchia, Judith Degen, Masoud Jasbi, Joshua Martin, Jack Rabinovitch, Giuseppe Riccardi, Uli Sauerland, Jesse Snedeker, and Julia Sturm for many helpful comments and insights along the way, and appreciate broader discussions with our audiences at the Harvard Meaning & Modality Lab, Harvard LangCog Workshop, Language Acquisition Lab at Tsinghua University, Experiments in Linguistic Meaning 2020, and Chicago Linguistic Society 2022. Y.Z. and K.D. acknowledge research funding from the Institute of Quantitative Social Sciences at Harvard University.

Competing interests

The authors have no competing interests to declare.

Authors' contributions

Both authors contributed to every aspect of the study, including but not limited to conceptualization of the study, experimentation with data curation and formal analysis, funding acquisition, and manuscript creation and revision. Y.Z. made the visualizations. K.D. supervised the overall study.

References

Ahn, D. (2020). *THAT thesis: A competition mechanism for anaphoric expressions* [Doctoral dissertation]. Harvard University. Retrieved August 16, 2023, from https://ling.auf.net/lingbuzz/004742

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*(3), 191–238. DOI: https://doi.org/10.1016/0010-0277(88)90020-0

Ambridge, B., Bidgood, A., Pine, J. M., Rowland, C. F., & Freudenthal, D. (2016). Is passive syntax semantically constrained? Evidence from adult grammaticality judgment and comprehension studies. *Cognitive Science*, *40*(6), 1435–1459. DOI: https://doi.org/10.1111/cogs.12277

Anderson, C. (2004). *The structure and real-time comprehension of quantifier scope ambiguity* [Doctoral dissertation]. Northwestern University. http://search.proquest.com.ezp-prod1.hul.harvard.edu/dissertations-theses/structure-real-time-comprehension-quantifier/docview/305136571/se-2

Anderson, J. C. (2013). Misreading like a lawyer: Cognitive bias in statutory interpretation. *Harvard Law Review*, *127*(6), 1–74. https://harvardlawreview.org/2014/04/misreading-like-alawyer/

Apperly, I. A. (2012). *Mindreaders: The cognitive basis of "theory of mind"* (1st ed.). Psychology Press.

Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A., & Moffitt, G. (2010). Why are there limits on theory of mind use? Evidence from adults' ability to follow instructions from an ignorant speaker. *Quarterly Journal of Experimental Psychology*, *63*(6), 1201–1217. DOI: https://doi.org/10.1080/17470210903281582

Apperly, I. A., & Robinson, E. J. (2003). When can children handle referential opacity? Evidence for systematic variation in 5- and 6-year-old children's reasoning about beliefs and belief reports. *Journal of Experimental Child Psychology*, *85*(4), 297–311. DOI: https://doi.org/10.1016/S0022-0965(03)00099-7

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37–46. DOI: https://doi.org/10.1016/0010-0277(85)90022-8

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. DOI: https://doi.org/10.1016/j.jml.2012.11.001

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387. DOI: https://doi.org/10.1016/0010-0285(86)90004-6

Bott, L., & Chemla, E. (2016). Shared and distinct mechanisms in deriving linguistic enrichment. *Journal of Memory and Language*, *91*, 117–140. DOI: https://doi.org/10.1016/j.jml.2016.04.004

Box, G. E., & Tiao, G. C. (2011). *Bayesian inference in statistical analysis*. John Wiley & Sons.

Brasoveanu, A., & Dotlacil, J. (2015). Sentence-internal same and its quantificational licensors: A new window into the processing of inverse scope. *Semantics and Pragmatics*, *8*(1). DOI: https://doi.org/10.3765/sp.8.1

Brasoveanu, A., & Dotlačil, J. (2019). Quantification. In C. Cummins & N. Katsos (Eds.), *The Oxford handbook of experimental semantics and pragmatics* (pp. 228–245). Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780198791768.013.3

Bürkner, P.-C. (2017). **brms**: An *R* package for Bayesian multilevel models using *Stan*. *Journal of Statistical Software*, *80*(1). DOI: https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395. DOI: https://doi.org/10.32614/RJ-2018-017

Capone, A. (2011). Default semantics and the architecture of the mind. *Journal of Pragmatics*, *43*(6), 1741–1754. DOI: https://doi.org/10.1016/j.pragma.2010.11.004

Charlow, S., & Sharvit, Y. (2014). Bound 'de re' pronouns and the LFs of attitude reports. *Semantics and Pragmatics*, *7*(3), 1–43. DOI: https://doi.org/10.3765/sp.7.3

Cohen, M., Tang, W., & Wang, Y. (2021). De re updates. *arXiv preprint arXiv:2106.11497*. DOI: https://doi.org/10.4204/EPTCS.335.9

Cremers, A., Fricke, L., & Onea, E. (2023). The importance of being earnest: How truth and evidence affect participants' judgments. *Glossa Psycholinguistics*, *2*(1). DOI: https://doi.org/10.5070/G6011172

Cummins, C., & Katsos, N. (Eds.). (2019). *The Oxford handbook of experimental semantics and pragmatics* (1st ed.). Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780198791768.013.33

Davidson, K. (2020). Is "experimental" a gradable predicate? *Proceedings of NELS 50*.

Deal, A. R. (2018). Compositional paths to de re. *Semantics and Linguistic Theory*, *28*, 622. DOI: https://doi.org/10.3765/salt.v28i0.4443

Degen, J., & Tanenhaus, M. K. (2016). Availability of alternatives and the processing of scalar implicatures: A visual world eye-tracking study. *Cognitive Science*, *40*(1), 172–201. DOI: https://doi.org/10.1111/cogs.12227

Donnellan, K. S. (1966). Reference and definite descriptions. *The Philosophical Review*, *75*(3), 281–304. DOI: https://doi.org/10.2307/2183143

Elliott, P. D. (2023). A flexible scope theory of intensionality. *Linguistics and Philosophy*, *46*(2), 333–378. DOI: https://doi.org/10.1007/s10988-022-09367-w

Epstein, S., Naumov, P., & Tao, J. (2023). An egocentric logic of de dicto and de re knowing who. *Journal of Logic and Computation*, Article exad053. DOI: https://doi.org/10.1093/logcom/exad053

Fodor, J. D. (1970). *The linguistic description of opaque contexts* [Doctoral dissertation]. Massachusetts Institute of Technology.

Fodor, J. D. (1982). The mental representation of quantifiers. *Processes, beliefs, and questions: Essays on formal semantics of natural language and natural language processing* (pp. 129–164). Springer. DOI: https://doi.org/10.1007/978-94-015-7668-0_5

Frege, G. (1948). Sense and Reference. *The Philosophical Review*, *57*(3), 209–230. DOI: https://doi.org/10.2307/2181485

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 457–472. DOI: https://doi.org/10.1214/ss/1177011136

Gibson, E., & Fedorenko, E. (2010). Weak quantitative standards in linguistics research. *Trends in Cognitive Sciences*, *14*(6), 233–234. DOI: https://doi.org/10.1016/j.tics.2010.03.005

Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, *16*(1), 1–14. DOI: https://doi.org/10.1017/S0140525X00028636

Jasbi, M., Bermudez, N., & Davidson, K. (2023). Default biases in the interpretation of English negation, conjunction, and disjunction. *Experiments in Linguistic Meaning*, *2*, 129. DOI: https://doi.org/10.3765/elm.2.5382

Jasbi, M., Bermudez, N., Zhang, Y., Siro, R., & Davidson, K. (2023). Crosslinguistic consistency in the interpretation of logical connectives: The case of English, Hungarian, Spanish, and Mandarin Chinese. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*(45). Retrieved August 7, 2023, from https://escholarship.org/uc/item/9tw5k7ff

Jasbi, M., Waldon, B., & Degen, J. (2019). Linking hypothesis and number of response options modulate inferred scalar implicature rate. *Frontiers in Psychology*, *10*, 189. DOI: https://doi.org/10.3389/fpsyg.2019.00189

Jaszczolt, K. (1997). The 'default de re' principle for the interpretation of belief utterances. *Journal of Pragmatics*, *28*(3), 315–336. DOI: https://doi.org/10.1016/S0378-2166(97)00006-4

Jaszczolt, K. (1999). *Discourse, beliefs and intentions: Semantic defaults and propositional attitude ascription*. Brill. Retrieved January 21, 2023, from https://brill.com/display/title/23328

Jaszczolt, K. (2005). *Default semantics: Foundations of a compositional theory of acts of communication*. Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199261987.001.0001

Jaszczolt, K. (2015). Default semantics. In B. Heine & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis* (2nd ed.). Oxford University Press.

Katsos, N., & Bishop, D. V. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*(1), 67–81. DOI: https://doi.org/10.1016/j.cognition.2011.02.015

Keshet, E. (2008). *Good intensions: Paving two roads to a theory of the de re/de dicto distinction* (Doctoral dissertation). Massachusetts Institute of Technology. Retrieved November 17, 2019, from https://dspace.mit.edu/handle/1721.1/45622

Keshet, E., & Schwarz, F. (2019). De re/de dicto. *The Oxford handbook of reference* (pp. 167–202). DOI: https://doi.org/10.1093/oxfordhb/9780199687305.013.10

Lederman, H. (2022). Fregeanism, sententialism, and scope. *Linguistics and Philosophy*, *45*(6), 1235–1275. DOI: https://doi.org/10.1007/s10988-022-09346-1

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means, aka least-squares means. *R package version*, *1*(1), 3. DOI: https://doi.org/10.32614/CRAN.package.emmeans

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001. DOI: https://doi.org/10.1016/j.jmva.2009.04.008

Lewis, S., Hacquard, V., & Lidz, J. (2017). "Think" pragmatically: Children's interpretation of belief reports. *Language Learning and Development*, *13*(4), 395–417. DOI: https://doi.org/10.1080/15475441.2017.1296768

Marty, P., Chemla, E., & Sprouse, J. (2020). The effect of three basic task features on the sensitivity of acceptability judgment tasks. *Glossa: A Journal of General Linguistics*, *5*(1), 72. DOI: https://doi.org/10.5334/gjgl.980

Nalborczyk, L., Batailler, C., Loevenbruck, H., Vilain, A., & Bürkner, P.-C. (2019). An introduction to bayesian multilevel models using brms: A case study of gender effects on vowel variability in standard indonesian. *Journal of Speech, Language, and Hearing Research*, *62*(5), 1225–1242. DOI: https://doi.org/10.1044/2018_JSLHR-S-18-0006

Nelson, M. (2019). The de re/de dicto distinction (Supplement to propositional attitude reports). In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2019). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2019/entries/prop-attitude-reports/dere.html

Nelson, M. (2023). Propositional attitude reports. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Spring 2023). Metaphysics Research Lab, Stanford University.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255–258. DOI: https://doi.org/10.1126/science.1107621

Paape, D., Vasishth, S., & Von Der Malsburg, T. (2020). Quadruplex negatio invertit? The online processing of depth charge sentences. *Journal of Semantics*, *37*(4), 509–555. DOI: https://doi.org/10.1093/jos/ffaa009

Partee, B. H., ter Meulen, A., & Wall, R. E. (1990). *Mathematical methods in linguistics*. Kluwer Academic.

Percus, O. (2000). Constraints on some other variables in syntax. *Natural Language Semantics*, *8*(3), 173–229. DOI: https://doi.org/10.1023/A:1011298526791

Percus, O., & Sauerland, U. (2003). On the LFs of attitude reports. *Proceedings of Sinn und Bedeutung*, 15. http://ling.uni-konstanz.de/pages/conferences/sub7/

Pullum, G. K. (2014). Fear and loathing of the English passive. *Language & Communication*, *37*, 60–74. DOI: https://doi.org/10.1016/j.langcom.2013.08.009

Quine, W. V. (1956). Quantifiers and propositional attitudes. *The Journal of Philosophy*, *53*(5), 177. DOI: https://doi.org/10.2307/2022451

Raffray, C. N., & Pickering, M. J. (2010). How do people construct logical form during language comprehension? *Psychological Science*, *21*(8), 1090–1097. DOI: https://doi.org/10.1177/0956797610375446

Rees, A., & Bott, L. (2018). The role of alternative salience in the derivation of scalar implicatures. *Cognition*, *176*, 1–14. DOI: https://doi.org/10.1016/j.cognition.2018.02.024

Richard, M. (1990). *Propositional attitudes: An essay on thoughts and how we ascribe them*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511625282

Roberts, C. (2012). Information structure: Towards an integrated formal theory of pragmatics. *Semantics and pragmatics*, *5*, 1–69. DOI: https://doi.org/10.3765/sp.5.6

Robinson, E. J., & Apperly, I. A. (2001). Children's difficulties with partial representations in ambiguous messages and referentially opaque contexts. *Cognitive Development*, *16*(1), 595–615. DOI: https://doi.org/10.1016/S0885-2014(00)00035-6

Romoli, J., & Sudo, Y. (2009). De re/de dicto ambiguity and presupposition projection. *Proceedings of Sinn und Bedeutung*, *13*(2), 14.

Ronai, E., & Xiang, M. (2021). Pragmatic inferences are QUD-sensitive: An experimental study. *Journal of Linguistics*, *57*(4), 841–870. DOI: https://doi.org/10.1017/S0022226720000389

Russell, B. (1905). On denoting. *Mind*, *14*(56), 479–493. http://www.jstor.org/stable/2248381. DOI: https://doi.org/10.1093/mind/XIV.4.479

Schwarz, F., Clifton, C., & Frazier, L. (2007). Strengthening 'or': Effects of focus and downward entailing contexts on scalar implicatures. *University of Massachusetts Occasional Papers in Linguistics*, *33*(1), 9. https://scholarworks.umass.edu/umop/vol33/iss1/9

Sison, C. P., & Glaz, J. (1995). Simultaneous confidence intervals and sample size determination for multinomial proportions. *Journal of the American Statistical Association*, *90*(429), 366–369. https://www.jstor.org/stable/2291162. DOI: https://doi.org/10.1080/01621459.1995.10476521

Sprouse, J., & Almeida, D. (2017). Design sensitivity and statistical power in acceptability judgment experiments. *Glossa: A Journal of General Linguistics*, *2*(1), 1–32. DOI: https://doi.org/10.5334/gjgl.236

Sudo, Y. (2014). On de re predicates. *Proceedings of WCCFL*, *31*, 447–456.

Tonhauser, J., Beaver, D. I., & Degen, J. (2018). How projective is projective content? Gradience in projectivity and at-issueness. *Journal of Semantics*, *35*(3), 495–542. DOI: https://doi.org/10.1093/jos/ffy007

Tonhauser, J., & Matthewson, L. (2015). *Empirical evidence in research on meaning*. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=c738bcc34514e95c4aaf7a745e14a7c08bc9b756

Tooley, K. M., & Traxler, M. J. (2010). Syntactic priming effects in comprehension: A critical review. *Language and Linguistics Compass*, *4*(10), 925–937. DOI: https://doi.org/10.1111/j.1749-818X.2010.00249.x

Tunstall, S. L. (1998). *The interpretation of quantifiers: Semantics and processing* [Doctoral dissertation]. University of Massachusetts, Amherst.

von Fintel, K., & Heim, I. (2011). *Intensional semantics*. http://lingphil.mit.edu/papers/heim/fintelheim-intensional.pdf

Wang, J. J., Ciranova, N., Woods, B., & Apperly, I. A. (2020). Why are listeners sometimes (but not always) egocentric? Making inferences about using others' perspective in referential communication (N. D. Duran, Ed.). *PLOS ONE*, *15*(10), Article e0240521. DOI: https://doi.org/10.1371/journal.pone.0240521

Wellman, H. M. (1992). *The child's theory of mind.* The MIT Press.

Wiebe, J., & Rapaport, W. (1986). Representing de re and de dicto belief reports in discourse and narrative. *Proceedings of the IEEE*, *74*(10), 1405–1413. DOI: https://doi.org/10.1109/PROC.1986.13641

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128. DOI: https://doi.org/10.1016/0010-0277(83)90004-5

Wu, Z., Merrill, W., Peng, H., Beltagy, I., & Smith, N. A. (2023). Transparency helps reveal when language models learn meaning. *Transactions of the Association for Computational Linguistics*, *11*, 617–634. https://aclanthology.org/2023.tacl-1.36. DOI: https://doi.org/10.1162/tacl_a_00565

Zhang, Y., & Davidson, K. (2021). De re interpretation in belief reports: An experimental investigation. *Experiments in Linguistic Meaning*, *1*, 310. DOI: https://doi.org/10.3765/elm.1.4874

Zhang, Y., Ricciardi, G., & Davidson, K. (2021). How many responses in a TVJT? It depends. *The 34nd Annual CUNY Conference on Human Sentence Processing*. https://doi.org/10.17605/OSF.IO/Z7M69

Zhang, Y., Ryskin, R., & Gibson, E. (2023). A noisy-channel approach to depth-charge illusions. *Cognition*, *232*, Article 105346. DOI: https://doi.org/10.1016/j.cognition.2022.105346