

De re interpretation in belief reports—An experimental investigation

Yuhan Zhang & Kathryn Davidson*

Abstract. Determiner phrases (DPs) under intensional operators give rise to multiple interpretations, known as the *de re/de dicto* ambiguity. Formal theoretical approaches to modeling this ambiguity must rely on nuanced semantic judgments, but inconsistent judgments in the literature suggest that informal judgment collection may be insufficient. In addition, little is known about how these ambiguities are resolved in context and how preferences between these readings vary by context and across individuals, etc. We reported three controlled experiments to systemize the truth-value judgment collection of *de re/de dicto* readings. While the *de dicto* readings were robustly accepted by nearly all English speakers, *de re* readings exhibited strongly bimodal judgments, suggesting an inherent disagreement among speakers. In addition, the acceptability of *de re* judgments was affected by the DP's internal structure as well as idiosyncratic scenarios. More broadly, our experimental results lend support to the practice of including quantitative data collection within semantics.

Keywords. *de re/de dicto* ambiguity; truth-value judgment; quantitative method

1. Introduction. In natural language semantics, the *de re/de dicto* ambiguity captures the referential properties of determiner phrases (DPs) in intensional domains¹. When the intensional operator is an attitude predicate, DPs read *de re* refer via description that holds in the context of the speaker; the attitude holder needs not themselves commit to the referential association between the linguistic expression and its referent. In contrast, DPs read *de dicto* have descriptions that need only hold in the possible worlds introduced by the intensional operator associated with the attitude holder, and consequently, the referential association might not hold for the speaker. For a concrete example, first consider the intensional predicate *want* and the indefinite DP *a prince*: in the sentence *Aurora wants to marry a prince* with the *de re* reading, the speaker communicates that there is a specific prince that Aurora wants to marry, even if she does not recognize him as a prince, while under *de dicto* reading, the speaker would be informing us that Aurora wants to marry whoever is a prince. (The first is true in the world of *Sleeping Beauty*; the second is false.) A (mostly, although not entirely) parallel situation happens with definite DPs: the sentence *Aurora wants to marry the prince*, under what we will call its *de re* reading, is true when Aurora's desires include marrying a specific person who is a prince in the

* Many thanks to Shannon Bryant, Judith Degen, Masoud Jasbi, Joshua Martin, Giuseppe Riccardi, Jack Robinovitch, Uli Sauerland, Jesse Snedeker, Julia Sturm, and audiences at the M&M Linguistics Laboratory, Harvard LangCog Workshop, and ELM 1 for critical feedback (mistakes are of course our own). Thanks to Yilan Wang for the assistance with the experiment pictures. Data collection was supported by a Graduate Student Research Grant from Institute for Quantitative Social Science at Harvard University awarded to YZ. Authors: Yuhan Zhang (yuz551@g.harvard.edu) and Kathryn Davidson (kathryndavidson@fas.harvard.edu), Harvard University.

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 adoption of the terms in philosophy and linguistics was initiated by Frege, Russell, and Quine but the current sense of *de re* and *de dicto* is not directly or intuitively related to the literal Latin meaning of the terminology (*de re*: 'of the thing', *de dicto*: 'of what is said') (von Fintel & Heim, 2011). Therefore, it is clearer to introduce the *de re/de dicto* distinction via contextualized examples. For more details of the nomenclature, see Keshet and Schwarz (2019).

actual world, even if she does not realize that he is the prince. What we will call here its *de dicto* reading is true when Aurora’s desires include “marrying the prince”, even if she is mistaken about who the prince may be.

In this paper, we examine current semantic judgment collection methods for formal theories on the *de re/de dicto* ambiguity. We report three novel experiments which highlight some advantages of obtaining quantitative judgments for readings with this ambiguity, focusing on definite DPs, given that they have already generated some judgment inconsistency in the literature. In the first section, we lay out the existing theoretical landscape and raise the motivation for the quantitative approach. In the second section, we report three controlled truth-value judgment experiments that tested whether the inconsistency also existed among multiple native English speakers. In the final section, we endeavor to account for the observed judgment variation.

1.1. A BRIEF OVERVIEW OF CURRENT THEORETICAL FRAMEWORKS. A traditional approach to the *de re/de dicto* distinction models it as a scope ambiguity: in the logical form *de re* (but not *de dicto*) DPs outscope the intensional operator to obtain the reading assigned in the actual world, as in (1) (see Russell 1905, Fodor 1970, Cresswell & von Stechow 1982 among others).

- (1) Aurora wants to marry a prince.
de re LF: $\exists x[\text{prince}_{w_0}(x) \wedge \forall w[\text{BEL}_{w_0}(\text{Aurora}, w) \rightarrow [\text{marry}_w(\text{Aurora})(x)]]]$
de dicto LF: $\forall w[\text{BEL}_{w_0}(\text{Aurora}, w) \rightarrow \exists x[\text{prince}_w(x) \wedge \text{marry}_w(\text{Aurora})(x)]]]$

While this theory can capture the distinction for simple (indefinite) DPs as in (1), when they appear under quantification the outscoped *de re* reading diverges nontrivially from the intended reading. In response, Percus (2000) proposes a solution using situation pronouns. Every verb phrase (VP) and noun phrase (NP) takes a situation pronoun as an unpronounced intensional variable that is bound by a higher lambda abstractor to get its world assignment. In this way, DPs can attain the *de re* interpretation via the logical binding of intensional variables while remaining *in-situ*. However, situation pronouns overgenerate, predicting readings not attested in natural language, resulting in a proposal by Keshet (2008, 2011) known as *split intensionality*, a return to the traditional scope-based theory with the employment of a type-shifting operator associated with an intensional operator. Raising a DP above this operator but below the intensional operator not only makes the DP an intensional argument, but also assigns it a different world index from the matrix clause. Consequently, DPs can be interpreted *de re* once they are raised above the type-shifting operator while remaining in the scope of the intensional operators. However, neither the scope-based theories nor the situation pronoun theory can adequately account for the “multiple-guise” scenario raised by Quine (1956), and consequently the theory of *de re* readings has been enriched by the addition of *concept generators* (see (4) below), permitting *de re* readings without movement (Percus & Sauerland 2003, Anand 2006, Charlow & Sharvit 2014).

The *de re/de dicto* distinction illustrates how formal theories evolve in response to new pieces of linguistic observation. Importantly, these linguistic observations not only include the well formedness of a sentence, but also its truth-value judgment offered by linguists given a corresponding scenario. Given the complexity of the data (especially the context/scenario!) involved in *de re/de dicto* judgements, we find it unsurprising that there is some judgment inconsistency from existing publications. We review some inconsistencies in the next section, and conclude that they call for a more consistent judgment collection approach, an approach that generate, in Tonhauser and Matthewson (2015)’s term, “stable, replicable, and transparent” judgments for observations of the same kind in the collection of such layered semantic data.

1.2. THE NEED FOR QUANTITATIVE RESEARCH. Inconsistencies found in linguistics and relevant fields constitute our motivation for the current study. In the linguistic theory literature, there exist some inconsistent *de re* judgments about the same DP structure in nearly identical linguistic environments. For instance, von Fintel and Heim (2011) employ sentence (2) to argue that the DP *your abstract* exhibits genuine ambiguity of *de re* and *de dicto* interpretations, while Nelson (2019) argues that the target DP *her brother* in (3), which has the same internal structure as *your abstract*, cannot be interpreted *de re*.

(2) John believes that [your abstract]_{*de re*} will be accepted.

de re truth condition:

John reviewed an amazing abstract and thought that it will be accepted. The speaker of this belief report has the additional knowledge that the abstract is written by the addressee “you” and thus utters (2), while John does not know the authorship of the abstract.

(von Fintel and Heim 2001:157)

(3) # Sally believes that [her brother]_{*de re*} is happy.

(Supposed) *de re* truth condition:

Sally hears a person laughing outside on the street who happens to be her brother. She believes that the person is happy, even though she does not recognize him as her brother.

(Nelson 2019:13)

Nelson’s reason to deny the *de re* reading is that the belief holder Sally does not conceive the person—Sally’s brother in real world—as her brother. Nelson takes Sally’s perspective and argues that *de re* should not be true given the scenario, while von Fintel and Heim claim that *de re* readings are a regular natural language phenomenon. One may wonder whether Nelson’s reasoning is representative of a dispreference for *de re*: to test this, it helps to gather judgment data across a wider variety of scenarios. Or, perhaps particular linguistic features, e.g., *your* vs. *her*, or the inanimate vs. animate possessives, contribute to the unexpected judgment inconsistency. Or, perhaps individuals vary in their preferred readings when faced with ambiguity. Controlled scenarios, examples, and a broader population allow us to test these hypotheses.

Further inconsistency in the linguistic literature arises in more complex cases. Charlow and Sharvit (2014) note one such disagreement: while they claim that the possessee *mother* in (4) should be *de dicto*, they report that another linguist who works on the same phenomenon finds it more natural under a *de re* reading.

(4) John believes that [every female student]_{*de re*} likes [her]_{*de re*} [mother]_{*de dicto*}.

LF: John believes- w_0

$[\lambda_8 \lambda_9 \lambda_1$ [every female student- w_0

$[\lambda_2[[G_8 t_2]-w_1$ likes- w_1 [G_9 her $_2$]- w_1 mother- w_1]]]]]

Truth condition of a “multiple-guise” scenario:

John comes into contact with every actual female student more than once, and each actual female student appears each time in a different guise. The same woman appears in two different guises and John fails to recognize this. He thinks he came into contact with two different women. Furthermore, in John’s mind, the mapping between the different guises is one-to-one. The sentence is about a specific scenario when John believes that *a* likes *b*’s mother, *c* likes *d*’s mother, and *e* likes *f*’s mother when in reality, $a = b$, $c = d$, and $e = f$.

Outside of theoretical linguistics, real-life instances provide data that are occasionally unexpected given theoretical predictions. While it is claimed that cardinal DPs cannot be interpreted *de re* (Musan 1995, Keshet 2008, Romoli & Sudo 2009), sentence (5), an utterance collected at an economic conference reported on *Language Log* by Liberman (2005), suggests otherwise.

(5) U.S. forces in Iraq have intentionally killed [12 journalists]*de re*.

de re truth condition:

There were 12 journalists killed by the U.S. forces in an attempt but the forces did not know the people they killed were journalists.

(Liberman on October 23, 2005)

Moreover, outside linguistics, researcher in related fields (i.e. philosophy, psychology, law) have claimed that the *de re* reading is easier to obtain in scenarios where both *de re* and *de dicto* are admitted, which is not (as far as we know) a claim that has been made within linguistics. In philosophy, Jaszczolt (1997) maintains that the directly referential property of definite noun phrases is more salient in communication and thus argues for a “default *de re* reading”. Her perspective finds its allies in cognitive science and developmental psychology. For example, when a participant and a character/protagonist in an experiment both know the identity of an object but the protagonist remains partially ignorant of the object’s certain properties, both children and adult participants fail to restrict their description to the properties already known by the protagonist to refer to the object when put into the protagonist’s shoes (Mitchell et al. 1996, Apperly & Robinson 2003, Apperly & Butterfill 2009, Low & Watts 2013). These observations suggest an egocentrism or reality bias explanation and the easiness of accessing information in actual reality but not others’ mental status may bias one to expect something like a “default *de re*” hypothesis. This bias is also bolstered in legal settings where the focus on a literal interpretation of defendant’s action rather than his intention to conduct such action during jury procedure echoes the “default *de re*” claim (Anderson 2013).

Despite the observed judgment inconsistency in linguistic literature and the “default *de re*” claim outside linguistics, there has been no experimental work that we are aware of that directly looks at the judgment preferences for *de re/de dicto* readings in a given scenario. While Hackl et al. (2009) has studied *transparent* versus *opaque* readings in intensional transitive predicates using online reading times and gathered evidence supporting the scope-based theory over the situation pronoun approach, their finding—QRed transparent DPs facilitate the processing of the following ACD site—does not directly address the judgment inconsistencies introduced above.

Fortunately, crowdsourcing techniques offer a systematic solution. By gathering offline judgments from multiple native speakers, we can detect whether the observed inconsistency results from idiosyncratic noises or inherent disagreement; by creating multiple scenarios to test a single phenomenon, we can confirm whether the judgment is robust to more variation. Crucially, by comparing the *de re* and *de dicto* readings of the same sentence under minimally different contexts, we can attempt a fair comparison of the judgment pattern and hopefully understand factors involved in different preferences for these readings. In sum, there is good reason to believe that the *de re/de dicto* literature can benefit from more quantitative methods.

1.3. RESEARCH GOAL. We aim to set up a simple and efficient experimental template for systematically obtaining stable, replicable, and transparent judgments of *de re/de dicto* readings across carefully controlled scenarios. We focus on *definite* DPs since several above examples

with questionable judgments are definite (e.g., *your abstract, her mother*), although we note that this definite non-*de re* reading differs from the traditional *de dicto* exhibited by indefinite DPs.

2. Experiment one. Exp.1 used highly controlled scenarios that permitted both *de re* and *de dicto* readings of definite DPs to probe its acceptability and preferences from native English speakers.

2.1. PARTICIPANTS. 120 adult native English speakers were recruited through Amazon’s Mechanical Turk. They received \$2 compensation for finishing the experiment.

2.2. MATERIALS AND DESIGN. In this experiment, participants read four written scenarios and gave their judgment of a target sentence based on each scenario (an example in Table 1²).

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.

JUDGMENT QUESTION

According to this story, please use the slider bar to indicate to what extent you agree or disagree with the following statement.

Target Sentence I:

Julie believes that **Elizabeth’s poem** will win the competition. (*de re*)

Target Sentence II:

Julie believes that **Nicole’s poem** will win the competition. (*de dicto*)



Table 1: Example scenario in Exp.1

In each scenario, there were two terms that described the target object (e.g., poem). The protagonist (e.g., *Julie*) associated one term *X* (e.g., *Nicole’s poem*) with the target object but in reality, *X* was not correct and the correct descriptive term *Y* (e.g., *Elizabeth’s poem*) was not known by the protagonist. If the wrong term *X* was used in reporting the protagonist’s belief, a *de dicto* reading would emerge; if the correct term *Y* was used a *de re* reading would emerge. Given this scenario, both readings were predicted to be true (e.g., Romoli & Sudo 2009).

In each scenario, the participants read either one of the two target sentences (Target Sentence I or Target Sentence II, varied between participants) and dragged a slider bar to show the extent they agree or disagree. After the participants’ decision, a numeric judgment score was

² The full experiment material and the statistical analysis in Exp.1-3 can be accessed via <https://osf.io/qgnr5/>. The online survey can be accessed via https://harvard.az1.qualtrics.com/jfe/form/SV_01UTaqO9hkkAcGF.

recorded (from “highly agree” = 100 to “highly disagree” = -100). Three sanity check sentences were additionally provided in each scenario—one was definitely correct, one was definitely wrong, and the last was uncertain. Successful judgments on these indicate the participants were attentive and thus eligible for inclusion in the data analysis. The advantage of a slider bar is its greater sensitivity to reveal potential judgments that would otherwise stay concealed due to the strong categorical implication in designs like the binary or the Likert scale (Marty et al. 2020).

Each participant read four scenarios each coupled with four sentences. Two of the four scenarios were randomly chosen for the *de re* condition and the other two for the *de dicto* condition. The participants were randomly assigned to one of six lists created for Latin Square design. The order of four stories was randomized, so was the order of the four sentences within each scenario. The entire survey was created on Qualtrics and distributed on Amazon’s Mechanical Turk.

2.3. RESULTS. We analyzed only the participants who correctly judged the correct and incorrect sanity checks at least 75% of the times. 115 participants (95.8%) were retained for the analysis.

The histogram in Figure 1 shows the judgment distribution of *de re/de dicto* sentences across all scenarios. While judgments for *de dicto* readings overwhelmingly aggregate toward the “highly agree” end, judgments for *de re* readings are bimodal—although more than half of the judgments are agreed with, another sizable proportion goes to the “highly disagree” edge.

We further analyzed the agreement proportion in each scenario, assuming it was appropriate to treat the continuous judgment as a binary variable given its categorical distribution. In Figure 2, the *de dicto* agreement rates are at ceiling for all scenarios while *de re* judgments exhibit larger variability across scenarios with a unanimously lowering effect ($\chi^2 = 79.13$, $df = 1$, $p < .001$).

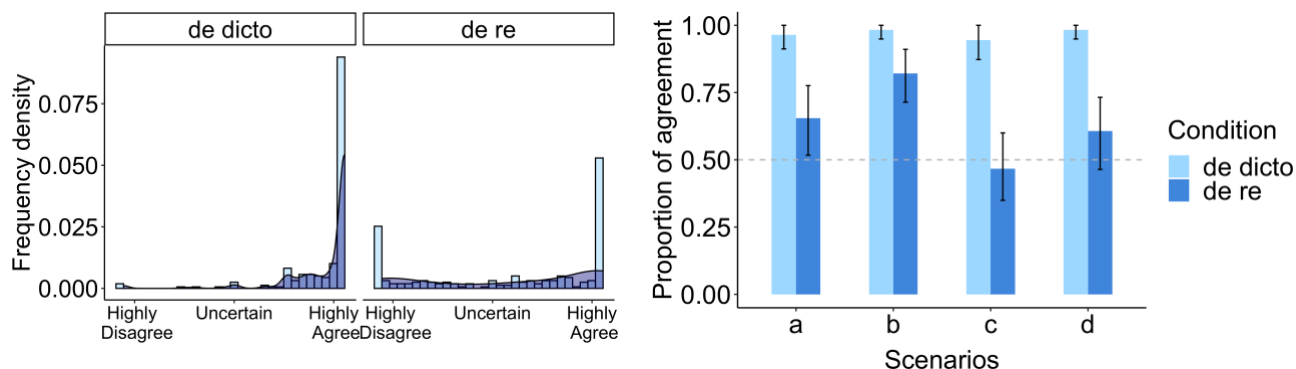


Figure 1 & 2: Histogram distribution of judgments; Agreement rate on condition and scenario³

The visual difference of *de re/de dicto* condition was confirmed by a mixed-effects logistic regression analysis. By treating the *de re/de dicto* conditions and the scenario as sum-encoded fixed effects with a random intercept on participants, we found that the *de re* trials were significantly less likely to be agreed with compared with a random trial ($\beta = -1.61$, $SE = 0.23$, $p < .001$); additionally, scenario B had a significantly higher agreement rate ($\beta = 0.96$, $SE = 0.30$, $p = .001$) while scenario C had a significantly lower one ($\beta = -0.81$, $SE = 0.24$, $p < .001$)⁴.

³ The error bars in Figure 2 represent 95% confidence intervals sampled via random bootstrapping.

⁴ The syntax of the model is “logit(agree) ~ condition + context + (1|subject)”. We also ran another model “logit(agree) ~ condition * context + (1|subject)” but didn’t find significant improvement from the reported one. The models were run using R with package *lme4*. The reason to treat scenario as a fixed effect rather than a random effect was that it was also of theoretical interests and that its number ($N = 4$) was not eligible for a random effect.

Additionally, we explored whether groups of participants had different judgment behavior, exhibited in Table 2. Clearly, a preponderance of participants agreed with both *de dicto* trials while the judgment behavior for *de re* had three representative groups, suggesting that an inherent disagreement among speakers or an unplanned scenario effect may both contribute to this distinctive participant behavior of interpreting *de re*.⁵

2.4. DISCUSSION. By setting up scenarios that theoretically allow both *de re* and *de dicto* readings of definite DPs and eliciting native speakers’ judgments, we found that while *de dicto* readings were unanimously available to participants, *de re* readings exhibited bimodal judgments with larger variations across scenarios and speakers. The sizable disagreement proportion and bimodal pattern suggests systematicity within the previously observed inconsistency in the literature.

# of Participants	Agree with 0 trial	Agree with 1 trial	Agree with 2 trials	Total
<i>de re</i>	21 (18.3%)	45 (36.5%)	52 (45.2%)	115
<i>de dicto</i>	0 (0.0%)	7 (6.1%)	108 (93.9%)	115

Table 2: Proportion of participants by the judgment behavior

3. Experiment two & three. While Exp.1 probed the judgement distribution for *de re* and *de dicto* reading of definite DPs (in particular, possessive constructions) in relatively simple sentences, Exp.2 and 3 continued asking whether the judgment disparity could extend to other DP structures or more sophisticated sentences. We were driven by this motivation to study the nuanced case of *bound de re* observed in Charlow and Sharvit (2014) for sentences like *John believes that every female student_i likes her_i mother* (above in (4)). The crucial *bound de re* assigns the QNP *every female student* and the possessive pronoun *her* a *de re* reading. The reading of *mother* is less critical for theoretical choices, but given these sentences are reported to have inconsistent judgments, we decided it was also worth investigating.

3.1. PARTICIPANTS. 160 participants in Exp.2 and 128 in Exp.3 took the tasks for \$2 on Amazon’s Mechanical Turk. After applying the same filter criterion as in Exp.1, 127 participants (79.38%) in Exp.2 and 120 (93.75%) in Exp.3 contributed to the analysis⁶.

3.2. METHODS. We treated the *de re/de dicto* reading of QNP as a between-experiments manipulation (*de re* in Exp.2 and *de dicto* in Exp.3) so that the possessive pronoun and the possessee, serving as two within-subjects manipulations, took either *de re/de dicto* reading within one experiment. Focusing on a 2 X 2 within-subjects manipulation within an experiment prevented participants from reading more than four complex scenarios and getting fatigued.

The general design is nearly the same as Exp.1. The most significant difference is that while Exp.1 had each scenario allow both *de re* and *de dicto* readings, Exp.2 and 3 were designed such that each scenario supported (e.g., made true) one reading, with the tested sentence being held constant within the experiment and the scenario being manipulated across conditions. There were also illustrative pictures to facilitate the processing (see an example in Table 3). Furthermore, Exp.2 and 3 had the same randomization, counterbalance, and filler design as Exp.1.

⁵ Thanks to Alexander Göbel for pointing to this inter-speaker investigation.

⁶ The lower inclusion rate in Exp.2 was because those participants were recruited on weekends when it is more challenging to gather good data via online implementation.

CONTEXT

As a photographer, John likes to rearrange his collections of photographs. One day, he encounters two sets of photos. In the first set, there are three ladies and each is holding a baby. John naturally believes that each of the babies is being held by their mother.



In the second set, three young adults are each wearing a T-shirt with a “2018” logo. John naturally believes that they were graduating students in the year 2018. John also notices that, interestingly, each of the young adults shares a similar smile to one of the ladies in the first photo set. He tries to recall if there is a connection between the young adults and the ladies but memory fails him.



As a matter of fact, what John fails to recall is three pieces of information. (1) The young adults in the second set of photos were the babies in the first set. They’ve grown up! (2) The ladies in the first set are actually the babies’ grandmother. The three young adults inherit their smile from their grandma who is mistakenly believed by John to be their mother. (3) The second set of photos were taken not in the graduation ceremony but when the three adults were volunteering for an academic conference in 2018.

Despite the fact that John doesn’t remember the correct relationship between the ladies in the first set of photos and the young adults in the second set and that he has incorrect information, John spends some time appreciating these photos.

JUDGMENT QUESTION

According to this story, please use the slider bar to indicate to what extent you agree or disagree with the following statement.

Target sentence: Looking at his photos, John believes that [every conference volunteer]_{de re} in the second set has the same smile as [their]_{de re} [mother]_{de dicto}.



Table 3: The canonical *bound de re* scenario, adapted from Charlow & Sharvit (2014)

3.3. RESULTS. Figure 3 shows that consistent with Exp.1, judgments tend to gather around both scale ends. Crucially, there is still a salient proportion of disagreement for the *bound de re* case compared with the “control” condition where all three nominal constructions were *de dicto*. Figure 4 shows the agreement rate of the eight conditions in eight columns. The second column

represents the canonical *bound de re* condition whose agreement rate is slightly above the chance level. Overall, the *de re* condition of the possessee leads to a lower agreement rate than the *de dicto* condition ($\chi^2 = 23.63$, $df = 1$, $p < .001$). Figure 5 displays the agreement rates by the condition manipulation and the scenarios. A visual examination shows a clear scenario effect because of the conspicuous lowering agreement rate in scenario four whose peculiarity was neither expected nor designed.

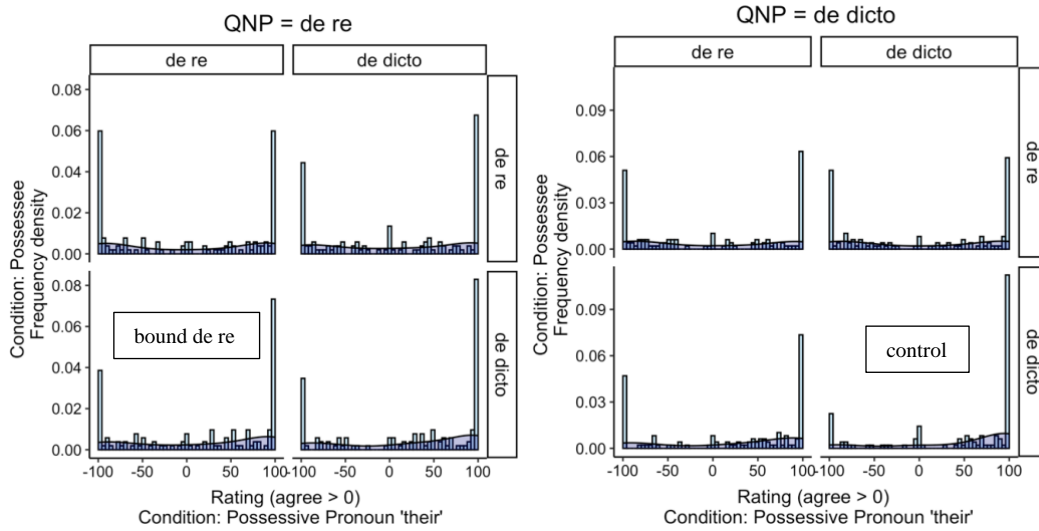


Figure 2: Agreement distribution in Exp.2 & 3

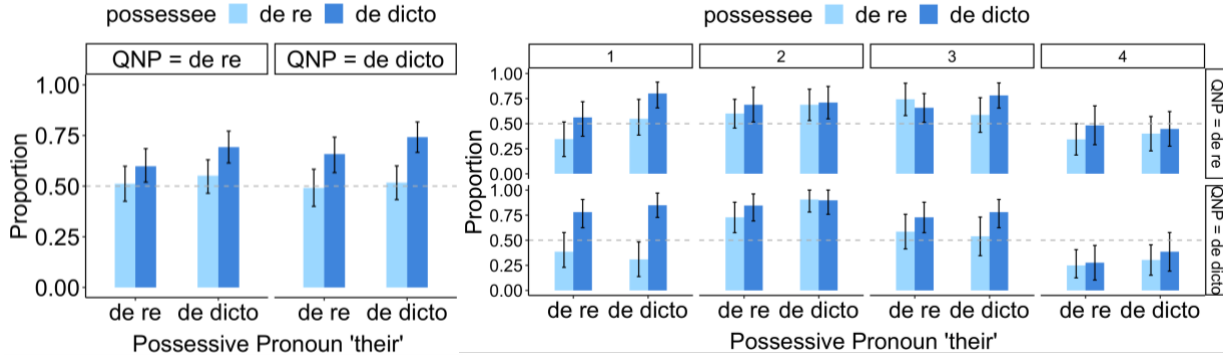


Figure 4 & 3: Agreement proportion among conditions (4); and scenarios (5)

The effects of *de re/de dicto* manipulation were further analyzed via a logistic mixed-effects model. The maximal model had one random intercept on participants and three fixed-effects variables to indicate the *de re/de dicto* assignment of the three nominal terms. The fourth fixed-effect variable was the story plus an interaction term between the story and each of the three nominal terms⁷. All the fixed effects were sum-coded.

The results show that while *de re* QNPs do not significantly affect the agreement rate ($\beta = -0.06$, $SE = 0.10$, $p = .561$), *de re* possessive pronouns ($\beta = -0.19$, $SE = 0.08$, $p = .02$) and *de re* possessee ($\beta = -0.42$, $SE = 0.08$, $p < .001$) did significantly lower the rate. Compared with the

⁷ The syntax of the model is “logit(agree) ~ (QNP + pronoun + possessee)*story + (1|subject)”. We’ve also fitted another model “logit(agree) ~ QNP + pronoun + possessee + story + (1|subject)” but the anova() test showed the former one outperformed the latter.

average agreement rate across the four scenarios and controlling the reading assignments of the three terms, scenario two ($\beta = 0.96$, $SE = 0.15$, $p < .001$) and three ($\beta = 0.41$, $SE = 0.14$, $p = .003$) were more likely to be agreed upon (more analysis on <https://osf.io/qgnr5/>).

3.4. DISCUSSION. Exp.2 and 3 tested the judgment of *de re* and *de dicto* in the *bound de re* type sentence and found that the canonical *bound de re* structure ([QNP]_{*de re*}, [Possessive pronoun]_{*de re*}, [Possessee]_{*de dicto*}) was agreed with more than 50% of the time, but at the same time (and, like *de re* readings generally) obtained salient disagreements. In general, Exp.2 and Exp.3 replicated the finding of Exp.1 in showing that *de re* readings lead to bimodal judgment. An additional result here is that the agreement rate of *de re* appears dependent on the DP's internal structure and/or position—the effect of *de re/de dicto* variation on the agreement rate was significant for the possessive pronouns and possesseees but nearly negligible for QNPs. Furthermore, the judgment was also affected by specific scenarios (e.g., peculiar Scenario 4).

Considering expectations of preferences between these readings, the salient difference in *de re* and *de dicto* readings of the possessee doesn't support the claim that a *de re* possessee is more natural, but rather is in line with Charlow and Sharvit's report. The lack of effect for QNPs echoes with numerous observations in theoretical work that both *de re* and *de dicto* readings for QNPs are felicitous (e.g., Mary 1978, Keshet 2008, Romoli & Sudo 2009 among others). Going back to the claim of *bound de re* in Charlow and Sharvit (2014), these two experiments suggest that (a) *bound de re* does exist for many speakers, but also that (b) not everyone agrees.

4. Conclusion and discussion. In a series of three experiments, we provided some quantitative evidence that there is a truth-value judgment disparity between *de re* and *de dicto* readings of definite DPs, both in simple attitude reports and more complex ones. In particular, we show that there is a systematic inconsistency for *de re* judgments: the bimodal distribution is far from a uniform or normal distribution we might expect if judgments were based on noise, and we note that this pattern of bimodal agreement would have been impossible to detect without quantitative methods that used response options more sensitive than a binary *true/false* (here, we used a slider bar). This inconsistency occurred not only in Exp.1 where the scenarios admitted both readings, but also in Exp.2 and 3 where the complex scenarios were controlled to admit only a single interpretation: even when it was the only one supported/true in the given scenario, *de re* readings had bimodal acceptance while *de dicto* readings were overwhelmingly accepted. Other relevant factors that also appear to affect the judgment include the internal structure of the DPs (e.g., possessive, quantificational, etc.) and features of the idiosyncratic scenarios.

While *de re* readings of combinations of other types of DP structures (including, crucially, indefinite DPs) and other kinds of attitude reports and intensional operators await testing, it is worth considering what may have led to disagreements in Exp.1 to 3. One cause might be that participants possess different grammars or dialects and one variation disallows the *de re* reading. This has to do with grammar variation and without information about the participants' linguistic profile, this claim will stay as a speculation. Another cause might be related to the scenario setup where the juxtaposition of *de re* and *de dicto* terms in the written scenario enhances participants' sensitivity to which term is used for reference in which possible world. The contrastive information evaluated in two parallel worlds could be well tracked by the participants and thus when their incremental comprehension starts from, for example, *Julie believes that...*, there is a chance that they only attend to what *Julie* believes and subsequently to descriptive terms held true in *Julie's* belief world. A *de dicto* DP naturally matches what the belief holder expects and thus is highly agreed upon, while encountering a *de re* DP whose referential relation to the entity is not held in *Julie's* mind could raise disagreement (as the case in example (3) raised by Nelson

(2019)). This sensitivity to the contrast also alludes to Theory of Mind ability and reasoning with perspective shifting (Wimmer & Perner 1983, Apperly & Butterfill 2009, Low & Watts 2013). For further investigation along these lines, it may be informative to experimentally test whether the contrast of *de re* and *de dicto* terms in the scenario influences *de re* judgment. That is, if there is no contrastive term such as *Nicole's poem* vs. *Elizabeth's poem* in one scenario but just one *de re* term unknown to the belief holder, what might be the *de re* agreement rate? If agreement increases, then we may conclude that the contrastive information may contribute to the high disagreement proportion of *de re*.

To conclude, our finding highlights the value of including quantitative methods as the basis for theoretical work, especially when the linguistic observation raises inconsistent judgments, as in the case of *de re* readings for definite DPs. The essential advantage of quantitative research is that with multiple speakers, multiple scenarios, and controlled manipulation, it is possible to detect whether an inconsistency observed from limited cases (e.g., sentences (2) and (3)) is underlyingly noise or concealed disagreement, and whether a preference for one reading is due to a minor contrast/preference or a grammatical unavailability. Here, we have uncovered evidence that inconsistencies about the *de re* readings were not due to noise or uncertainty among participants (which would lead to more intermediate agreement responses) but rather to systematic bimodal judgments and affected by scenario and DP-specific factors. We speculate that our finding may be especially enlightening in the case of known semantic/scope ambiguities, where naive speakers' preferences may affect their judgment behavior. What, then, to do with such results is an important question for theoreticians. Lastly, we end by noting that these results did not even require the researchers to create a massive number of scenarios to uncover these patterns: in our Exp1. to Exp.3, a mere four scenarios—just “a little bit experimental” in Davidson (2020)'s term—were enough to observe this systematic disagreement, which held across each of the experiments. We hope that this work will lead not just to more work along the *de re/de dicto* line, but contribute to the growing field of Experiments in Linguistic Meaning (ELM)!

References

- Anand, Pranav. 2006. *De de se*. PhD Thesis, Massachusetts Institute of Technology.
- Anderson, Jill C. 2013. Misreading like a lawyer: Cognitive bias in statutory interpretation. *Harvard Law Review* 127(6). 1–74.
- Apperly, Ian A. & Butterfill, Stephen A. 2009. Do humans have two systems to track beliefs and belief-like states? *Psychological Review* 116(4). 953–970. <https://doi.org/10.1037/a0016923>
- Apperly, Ian 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. [https://doi.org/10.1016/S0022-0965\(03\)00099-7](https://doi.org/10.1016/S0022-0965(03)00099-7)
- Charlow, Simon & Sharvit, Yael. 2014. Bound “*de re*” pronouns and the LFs of attitude reports. *Semantics and Pragmatics* 7(3). 1–43. <https://doi.org/10.3765/sp.7.3>
- Cresswell, Maxwell J. & von Stechow, Arnim. 1982. *De re* belief generalized. *Linguistics and Philosophy* 5(4). 503–535. <https://doi.org/10.1007/BF00355585>
- Davidson, Kathryn. 2020. Is “experimental” a gradable predicate? *Proceedings of North East Linguistics Society (NELS)* 50.
- Fodor, Janet D. 1970. *The Linguistic Description of Opaque Contexts*. PhD Thesis, Massachusetts Institute of Technology.

- Hackle, Martin, Koster-Moeller, Jorie, & Gottstein, Andrea. 2009. Processing opacity. *Proceedings of Sinn und Bedeutung* 13. 171–185.
- Jaszczolt, Kasia M. 1997. The ‘default *de re*’ principle for the interpretation of belief utterances. *Journal of Pragmatics* 28(3). 315–336. [https://doi.org/10.1016/S0378-2166\(97\)00006-4](https://doi.org/10.1016/S0378-2166(97)00006-4)
- Keshet, Ezra. 2008. *Good intensions: Paving two roads to a theory of the de re / de dicto distinction*. PhD Thesis, Massachusetts Institute of Technology.
- Keshet, Ezra. 2011. Split intensionality: A new scope theory of *de re* and *de dicto*. *Linguistics and Philosophy* 33(4). 251–283. <https://doi.org/10.1007/s10988-011-9081-x>
- Keshet, Ezra & Schwarz, Florian. 2019. De Re/De Dicto. In Jeanette Gundel & Barbara Abbott (eds.), *The Oxford Handbook of Reference*, 167–202. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199687305.013.10>
- Lieberman, Mark. 2005. Rarely better than *de re*. Blog. *Language Log*. <http://itre.cis.upenn.edu/~myl/languageblog/archives/002573.html>
- Low, Jason & Watts, Joseph. 2013. Attributing false beliefs about object identity reveals a signature blind spot in humans’ efficient mind-reading system. *Psychological Science* 24(3). 305–311. <https://doi.org/10.1177/0956797612451469>
- Marty, Paul, Chemla, Emmanuel, & Sprouse, Jon. 2020. The effect of three basic task features on the sensitivity of acceptability judgment tasks. *Glossa: A Journal of General Linguistics* 5(1). 72.1–23. <https://doi.org/10.5334/gjgl.980>
- May, Robert C. 1978. *The grammar of quantification*. PhD Thesis, Massachusetts Institute of Technology.
- Mitchell, P., Robinson, E. J., Isaacs, J. E., & Nye, R. M. 1996. Contamination in reasoning about false belief: An instance of realist bias in adults but not children. *Cognition* 59(1). 1–21. [https://doi.org/10.1016/0010-0277\(95\)00683-4](https://doi.org/10.1016/0010-0277(95)00683-4)
- Musan, Renate. 1995. *On the temporal interpretation of noun phrases*. PhD Thesis, Massachusetts Institute of Technology.
- Nelson, Michael. 2019. The *de re/de dicto* distinction (Supplement to Propositional Attitude Reports). In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/spr2019/entries/propositional-attitude-reports/dere.html>
- Percus, Orin. 2000. Constraints on some other variables in syntax. *Natural Language Semantics* 8(3). 173–229. <https://doi.org/10.1023/A:1011298526791>
- Percus, Orin & Sauerland, Uli. 2003. On the LFs of attitude reports. *Proceedings of Sinn und Bedeutung* 7. 228–242.
- Quine, Willard V. 1956. Quantifiers and Propositional Attitudes. *The Journal of Philosophy* 53(5). 177–187. <https://doi.org/10.2307/2022451>
- Romoli, Jacopo & Sudo, Yasutada. 2009. *De de/de dicto* ambiguity and presupposition projection. *Proceedings of Sinn und Bedeutung*, 13.
- Russell, Bertrand. 1905. On Denoting. *Mind* 14(56). 479–493.
- Tonhauser, Judith & Matthewson, Lisa. 2015. Empirical evidence in research on meaning. *Manuscript*.
- von Fintel, Kai & Heim, Irene. 2011. *Intensional Semantics*. *Manuscript*. <http://lingphil.mit.edu/papers/heim/fintel-heim-intensional.pdf>
- Wimmer, Heinz & Perner, Josef. 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children’s understanding of deception. *Cognition* 13. 103–128.