Do Lawyers Misread the Law? *De Re* and *De Dicto* Interpretation in Belief Reports

Joshua Boivin
Queen Mary, University of London

June 2025

Abstract

This study explores the accessibility of de re and de dicto interpretations in belief reports, with a focus on whether legal training affects these interpretations. Prior research suggests that lawyers, due to their specialized training, process such ambiguities differently from non-lawyers (Anderson 2014). To investigate this, I replicated and extended Zhang and Davidson's (2021, 2024) truth-value judgment task, which tested the acceptability of de re and de dicto in contexts where both are theoretically true. Notably, to probe whether either de re or de dicto is more accessible, response duration was recorded. The results of the judgment task were analyzed using a Bayesian multilevel model and an ordinal regression model while the response duration data was analyzed using a linear mixed effects model. The findings suggest that, while both groups show consistent agreement with de dicto interpretations, there is significant variability in de re judgments. Nevertheless, no evidence was found that either reading is more accessible, since not all contexts showed a difference between the two conditions and response duration was not shorter for de dicto than de re trials. Furthermore, the differences between lawyers and non-lawyers were minimal, suggesting that legal training does not significantly alter the processing of the de re/de dicto ambiguity. However, individual differences and context-dependent factors were shown to likely influence how participants process the belief reports.

Keywords belief reports, de re/de dicto ambiguity, experimental semantics, language and law, truth value judgment task

Queen Mary's OPAL #53
Occasional Papers Advancing Linguistics

[*] My thanks go to Dr Linnaea Stockall, Rémi Lamarque, and the audiences of the London Semantics Day and the BerlinBrnoVienna Workshop.

1 Introduction

In the 1970s David Smith moved into a flat in London. To install a new stereo system, he built in new electrical wiring and covered it with new flooring and wall paneling. When he moved out he removed the flooring and paneling to retrieve the wiring and take the stereo with him. As Rodes (1998) explains, this would have been unproblematic except that under the law of fixtures the new flooring and paneling had become the landlord's property and according to section 1 of the Criminal Damage Act of 1971 "A person who without lawful excuse destroys or damages any property belonging to another intending to destroy or damage any such property [...] shall be guilty of an offence."

It is clear, therefore, that there was a criminal act (*actus reus*): Smith destroyed his landlord's property. Since Smith intentionally destroyed property that did not belong to him (irrespective of the fact that he believed it was his own property), he was found guilty at the first instance. In other words, Smith was convicted on a *de re* interpretation of the Criminal Damage Act, since there was property which was not his own and he destroyed this property intentionally. The court of appeal opted for a different reading of the law. Smith was acquitted on a *de dicto* interpretation at this higher instance; for how could he intentionally damage someone else's property if he believed it was his own (cf. Regina v. Smith (David) [1974])? Put differently, Smith had no criminal intention (*mens rea*) and, as is the case for most criminal offences, Smith could only be convicted if there was both a criminal act and a criminal intention (cf. ICLR 2024, Regina v. Smith (David) [1974]).

(1) Smith believed that he destroyed his landlord's property

Simply put, the critical question in this case is whether (1) is true or false. At first instance (1) was found to be true and Smith was convicted. By contrast, the court of appeal found (1) to be false resulting in Smith's acquittal. Of course, both interpretations of (1) are possible and are an example of the *de re*/*de dicto* ambiguity. *De re* and *de dicto* are two different readings in ambiguous sentences which arise with adverbs like *intentionally* or *deliberately* and more notably with verbs like *believe*, *think*, or *intend* which describe mental states.

The classical approach to describe the *de re*/*de dicto* ambiguity is a difference in quantifier scope (Russell 1905, Fodor 1970, Cresswell and von Stechow 1982, Romoli and Sudo 2009 among others). To illustrate, (2) shows two possible logical forms for (1). In (2-a), the logical form for the *de re* reading, the universal quantifier is in the existential quantifier's scope, whereas in (2-b), the *de dicto* reading, it is the other way round. This means that in (2-a) the existence of something being the landlord's property in the utterance world ("the world we are in"; indicated by w*) is established first and it is then stated that in all possible worlds which Smith believes to be true (stated as DOX_{w*}), Smith destroyed this property. In (2-b) on the other hand, it is stated that in all possible worlds which Smith believes to be true there exists some property which belongs to the landlord and Smith destroyed this property. It should become clear that for (2-a) to be true it is not necessary that Smith believes that the property belongs to the landlord, while for (2-b) to be true it is.

(2) a. $\exists$x[landlord's property_{w*}(x) $\wedge \forall$w [DOX_{w*}(Smith, x) $\rightarrow$ [destroy_{w}(Smith)(x)]]]

DE RE

b. $\forall$w[DOX_{w*}(Smith, w) $\rightarrow \exists$x[landlord's property_{w}(x) $\wedge$ destroy_{w}(Smith)(landlord's property)]]

DE DICTO

So, in Smith's case it was ultimately acknowledged (though not explicitly) that both a *de re* and a *de dicto* interpretation of the law exist based on which Smith was at first found guilty and later acquitted.[1] However, this is not always the case. According to Anderson (2014) people with legal training (I will refer to them as "lawyers") are *de re* biased. She argues that lawyers often neglect the possibility of a *de dicto* interpretation which leads to verdicts that seem to contrast with the intention behind the laws. Evidently, this is not always true. Therefore, this

[1]More specifically, in Regina v. Smith (David) [1974] the question addressed was whether *mens rea* had to apply to both the property and the act or just the act.

dissertation's main goal is to answer the question whether lawyers do in fact have a *de re* bias. Based on Anderson (2014) it is expected that lawyers show a higher acceptance of *de re* than non-lawyers and that lawyers are more likely to reject *de dicto* while non-lawyers accept it throughout.

In a truth-value judgment task, Zhang and Davidson (2021, 2024) investigated the acceptability of *de re* and *de dicto* statements in contexts where both are theoretically true. They found that *de dicto* was accepted as true consistently while *de re* exhibited a bi-modal distribution. Zhang and Davidson make no claims about the accessibility of *de re* and *de dicto*, however, these results seemingly suggest that *de dicto* is more easily accessed than *de re* in the general population. To answer the question whether the *de re* bias among lawyers exists, Zhang and Davidson's experiment was replicated for this dissertation with the addition of a comparison between lawyers and non-lawyers. Furthermore, response duration was collected as a potential measure of accessibility. Previous literature has claimed that *de re* is the more accessible reading in *de re*/*de dicto* ambiguous statements. Due to fallacies in some of this literature (see section 2.1) and Zhang and Davidson's results, it is expected that *de dicto* responses will be formulated quicker than *de re*. This would be further evidence that *de dicto* is more accessible than *de re*.

In section 2.1, I will explore whether there is evidence for the fact that one of the two readings in the *de re*/*de dicto* ambiguity is more accessible than the other. This paves the way for the discussion of the literature concerned with *de re* and *de dicto* in law in section 2.2. At this point, in section 2.3, I will state the research questions and hypotheses more precisely. I will then present the methods used to answer these questions in section 3, followed by the results in section 4. These will then be discussed in section 5, before I end on some concluding remarks.

2 Background

2.1 *De Re*/*De Dicto*: Is one more accessible?

Different approaches in the formal semantics literature try to account for the *de re*/*de dicto* ambiguity. The classical approach, scope theory, does not predict either *de re* or *de dicto* to be more accessible. Other accounts of the *de re*/*de dicto* ambiguity are the intensionable variable theory and the presupposition projection theory (see Romoli and Sudo 2009 for a succinct overview). These theories do not make any claims about whether *de re* or *de dicto* should be more accessible either (Zhang and Davidson 2024). Hence, whether one reading can be derived from the other is not part of any of these approaches.

Nevertheless, there has been some debate on whether one of the two readings in *de re*/*de dicto* ambiguous sentences is more accessible than the other.[2] Despite the fact that, to my knowledge, there is no empirical evidence from linguistics that *de re* should be more accessible than *de dicto*, it has been claimed that this should be the case. For example, in Default Semantics, Jaszczolt (1997) and Capone (2011) have suggested that *de re* should be a default, whereas *de dicto* is the alternative.

Default Semantics, as the name suggests, assumes a set of default interpretations which are understood by the addressee automatically unless given a reason to opt for an alternative interpretation (see Jaszczolt 2015 for a more detailed and extensive description of Default Semantics). Jaszczolt's approach does not rely on a particular account of the *de re*/*de dicto* ambiguity but rather adds the mechanism of Default Semantics to a description of the ambiguity. *De re* is the default and *de dicto* the alternative based on the claim that *de re* is more accessible.[3] According to Jaszczolt (1997), the goal of a conversation is for the addressee to understand the speaker which is done by identifying the referents of the utterance. Since *de re* refers to a specific object, it is easier to identify the reference than *de dicto* which refers to anything matching the description. The idea is that *de re* highlights the referential property of the determiner phrase (the landlord's property in our example from above) because it refers to one object in the utterance world and not to a description in any possible world. In other words, it is easier to access

[2]In this dissertation, I will focus on what Zhang and Davidson (2024:12) call the "referential *de re*" and "referential *de dicto*" reading of sentences like (1) which are the two readings discussed so far. I will sideline the discussion of what they label as the "attributive *de dicto*" reading which would be true in Smith's case if he had no specific property in mind but believed that he destroyed whichever property belongs to his landlord (Zhang and Davidson 2024:12)

[3]To facilitate the understanding of this approach, here is another way to describe *de re* and *de dicto*: *De re* is about a specific thing (while it does not matter how this thing is referred to), whereas *de dicto* is about a description and whatever fits this description (Jaszczolt 1997:317). So bringing it back to Smith's case from above, *de re* is about the landlord's property (and it does not matter whether Smith is aware that it is the landlord's property), whereas *de dicto* is about the thing that is called the landlord's property.

something (a specific object) in the utterance world than something in somebody's belief world (a description which is mapped onto a specific object in each world).

It may seem plausible that a specific object can be more easily accessed than anything that fits a certain description, since there is one object in one world in the former and one object in many worlds in the latter case. However, this approach ignores the fact that it may be harder to identify the referent in this one specific world than to identify many objects in different belief worlds. Simply put, it may be harder to access the one utterance world than all of the belief worlds somebody holds. A simple example of this could be a case where the object being referred to does not actually exist in the real world. For example, in *Mary believes that a unicorn hit John*, it is arguably not easy to identify the unicorn in the real world, since it does not exist. By contrast, many objects that match the description of a unicorn in any possible belief worlds should be easier to access, since in these belief worlds an object matching the description of *unicorn* may very well exist. To my knowledge there is no empirical evidence that shows whether identifying referents is easier in the utterance world or belief worlds.

Theory of Mind experiments have contributed to some linguists thinking that *de re* should be more accessible than *de dicto*. Many studies have shown that children struggle when reasoning about mental states (Birch and Bloom 2004, Roeper 2007 among others). In a typical experiment a child is shown a picture of a person putting an object into one of two boxes. This person then leaves and the child is shown a picture of a second person moving the object into the other box. When asked where the first person thinks the object is, the child typically answers that the person thinks the object is in the second box. Of course this is not correct and the explanation is that the child cannot differentiate what other people know from what it knows itself. Adults are said to be able to keep apart what they know to what somebody else knows, but struggle when it is embedded as in *John thinks that Mary knows that Paul robbed the bank* (Kinderman et al. 1998, Rutherford 2004). Birch and Bloom (2007) also showed that with simple false beliefs adults show biases in their expectation of how a person with such a false belief will act. And Apperly et al. (2008) showed that adults processed slower and were more prone to error when informed about a false belief which conflicted with their own belief.

Anderson (2014) and Zhang and Davidson (2021) claim that these results show that *de re* is more accessible than *de dicto*. This seems plausible at first, however, they make a critical mistake. As shown by w* in the logical forms in (2), the difference between *de re* and *de dicto* is that *de re* is about something the way it is in the utterance world (or "in reality") and *de dicto* is about something the way it is in somebody's belief worlds. Although many Theory of Mind experiments claim to show that children and adults struggle more with somebody else's beliefs than with reality, they actually show that children and adults struggle more with someone else's beliefs than their own. In these experiments the participants beliefs often coincide with the reality constructed in the experiment but while we may hope that the world the way we think it is, is the utterance world, it is in no way certain that this is ever the case.

(3) a. CONTEXT: 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 (3-b), while John does not know the authorship of the abstract.

(Zhang and Davidson 2021:2)

b. John believes that [your abstract]$_{\text{DE RE}}$ will be accepted

(von Fintel and Heim 2011:157)

(4) a. CONTEXT: 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.

(Zhang and Davidson 2021:3)

b. #Sally believes that [her brother]$_{\text{DE RE}}$ is happy

(Nelson 2019:13)

So far I have discussed accounts that claim *de re* should be more accessible than *de dicto*. A first clue that *de re* may actually be less accessible than *de dicto* is that there is a dispute in the formal literature on the availability of *de re* readings in specific constructions. More specifically, von Fintel and Heim (2011) say that a *de re* reading of (3-b) is possible. The context provided in (3-a) shows under which conditions (3-b) should be true. So, in a situation where John is unaware of the authorship of the abstract but the speaker knows that it was written by the addressee, if the speaker knows that John thinks the addressee's abstract will be accepted, this sentence should be true. Nelson (2019), on the other hand, believes that (4-b) cannot be understood *de re*, but only *de dicto*. The sentence has the same structure and the DP is in a possessive construction in both cases. If (4-b) is deemed to be a possible sentence in a context like (4-a), Nelson should be wrong. It does not matter whether (4-b) is possible or not in (4-a), but the mere fact that there is a disparity among formal linguists is a clue that *de re* may not be that easily accessed. Furthermore, no such disagreement about the availability of *de dicto* exists in the literature to my knowledge.

(5) a. 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.

b. Julie believes that Elizabeth's poem will win the competition DE RE

c. Julie believes that Nicole's poem will win the competition DE DICTO

(Zhang and Davidson 2021:314)

Empirical evidence for a more accessible *de dicto* comes from experimental linguistics. Zhang and Davidson (2021) conducted a study in which participants performed a simple truth-value judgment task in which they accepted or rejected *de re* and *de dicto* statements in contexts in which both are theoretically true. For example, in (5-a) the *de re* statement, given in in (5-b), is *de re* in the sense that it is true on a *de re* reading. In the same way, (5-c) is *de dicto* in the sense that this sentence is true on its *de dicto* reading. 120 native English speakers read four short stories and then judged the acceptability of four sentences regarding each story context. For each story there was one critical trial. The other three sentences were statements that could be clearly judged and ensured that participants were attentative. Every participant judged two *de re* and two *de dicto* statements. The participants used a slider bar to indicate to what extent they agreed, felt uncertain or disagreed with each trial. The use of a slider bar allowed for more nuance than a binary agree or disagree.

Zhang and Davidson (2021:315) found that *de dicto* is consistently accepted in all cases, while *de re* shows a bi-modal distribution, with 50% of true *de re* statements accepted, but 25% rejected. Moreover, *de dicto* is also highly accepted across all scenarios while *de re* has much lower agreement rates ranging from less than 50% for scenario C, up to over 75% for scenario B. *De re* was shown to be significantly less likely to be agreed with by participants than a random trial. Furthermore, 94% of participants agreed with both *de dicto* trials and 0% with neither of them compared to only 45% of participants who agreed with both, 37% who agreed with one, and 18% with neither of the *de re* trials. Nelson and von Fintel and Heim's disagreement on the availability of a *de re* reading from above could plausibly be a reflection of the results of this study (Zhang and Davidson 2021).

This study was replicated by Zhang and Davidson (2024) with 60 monolingual English speakers from the USA with the minor adjustment that a Likert scale was used instead of a slider bar which preserved the benefits over a binary scale while facilitating the analysis of the results. They found a high acceptance of *de dicto* and a bimodal distribution for *de re* again. However, they found variation in the judgment distribution across the four contexts. Fitted into a Bayesian model, Zhang and Davidson reported no difference between *de dicto* in contexts A (β = 1.40, HPD = [-0.16, 2.99]) and B (β = -1.30, HPD = [-3.84, 0.95]). By contrast, they found a difference in contexts C (β = 3.48, HPD = [1.33, 5.98]) and D (β = 3.10, HPD = [0.65, 5.68]). While *de dicto* was agreed with consistently, "Highly Agree" reaching around 65% in contexts A, B, and C and ca 85% in D, the proportion of "Highly Agree" in the *de re* condition stayed below 50% in all contexts except B (ca 75%) (Zhang and Davidson 2024:19).

The fact that overall *de dicto* was robustly accepted while *de re* was not, suggests that *de dicto* may be more accessible than *de re*. Moreover, that both *de dicto* trials were accepted by almost all participants while more than half did not accept both *de re* trials is further evidence that *de re* is less accessible than *de dicto*.

In another experiment, Zhang and Davidson (2024:27) found that in contexts where only *de re* statements can be true *de re* was highly accepted by the participants. A context that only allows *de re* to be true differs from the one given in (5-a) in that Julie is not falsely assuming the authorship but is simply completely ignorant towards who may be the author. In a context like the above, where both *de re* and *de dicto* are considered to be true by the formal semantics literature, the bimodal distribution was observed again. The fact that *de re* becomes highly accepted in a context where it is the only possible option shows that *de re* is available to speakers. The fact that in a context where *de re* is in competition with *de dicto*, *de re*'s acceptance dwindles, is again evidence for the fact that *de dicto* is more accessible than *de re*.

While Zhang and Davidson's (2021) results were replicated in 2024, they should be regarded with caution in light of the so-called "replication crisis" which engulfed the social sciences since the late 2000s (Shrout and Rodgers 2018). Sönning and Werner (2021) highlight that there has been a trend across all fields of linguistics for the past two decades towards more quantitative methods. They stress that linguistics, which is comparatively new to the field, has a special need for replication studies. Zhang and Davidson's studies can be seen as an example of this shift from qualitative and theoretical to quantitative methods. Therefore, the call for caution is even more important. Since both of Zhang and Davidson's studies were conducted in the USA, it is important to replicate their experiment with a different population.

To sum up, the evidence for *de re* being more accessible is unconvincing. The experimental data presented from linguistics is stronger evidence for the fact that *de dicto* is more easily accessed than *de re*. This is manifested in that *de dicto* is more widely accepted by speakers than *de re* in contexts permitting both, and that a *de dicto* and *de re* permitting context decreases *de re* acceptance compared to *de re*-only permitting contexts.

2.2 Lawyers and the *De Re*/*De Dicto* Ambiguity

Despite literature suggesting that it is more plausible that the general population accesses *de dicto* more easily than *de re*, Anderson (2014) claims that lawyers are *de re* biased. She goes as far as saying that lawyers misread in statutory interpretation. Specifically, she argues that they overlook *de dicto* readings of ambiguously written laws. In four cases covering fraud by impersonation, obstruction of justice, genocide, and disability rights, Anderson illustrates how courts reached verdicts based solely on *de re* interpretations which clearly go against the intention of the actions the law was supposed to prevent. A *de dicto* interpretation would have captured this in these four cases.

For example, in the fraud by impersonation case, a man went to vote as his neighbor because, in contrast to his neighbor, he had no right to vote. However, the neighbor he was impersonating was dead. Since the relevant law forbade impersonating any person entitled to vote , the defense argued that the defendant hadn't broken the law because he impersonated a dead person and a dead person is not entitled to vote – the court begrudgingly agreed (Whiteley v. Chappell [1868]). This is of course a *de re* interpretation of the law where the person entitled to vote must exist in reality. On a *de dicto* reading the defendant could have been convicted, since arguably the law meant to stop ineligible people from voting. The law was eventually expanded to explicitly forbid impersonating not only alive but also dead and fictitious people. Simply using a *de dicto* interpretation would have made this amendment superfluous (Anderson 2014).

Anderson notes that not only lawyers are *de re* biased. She compares lawyers to children and people with autism who have been said to struggle with *de dicto* readings because *de dicto* requires thinking about others' belief worlds rather than the real world. However, to my knowledge there are no linguistic studies on the acquisition of *de re* and *de dicto*, but merely conclusions drawn from Theory of Mind research which predict *de re* to be acquired by children before *de dicto*. In section 2.1 I addressed why this evidence is not convincing. Anderson further builds an argument that *de re* must be more accessible than *de dicto*. Yet, she concedes that *de dicto* does not pose any problems for communication among adults in everyday conversation. She links this to Kahneman's (2012) idea of thinking fast and slow.

Since *de re* is more accessible, according to Anderson this should be accessible fast, whereas *de dicto* is slow. As she states herself, lawyers are of course actually thinking slow and taking a lot of time in statutory interpretation, trying to find an angle or a way to use the law to benefit their case. So, how come lawyers become oblivious to *de dicto*? Anderson says that lawyers are trained to anchor the facts in the real world which explains the preference for *de re*. She suggests that somehow their very slow thinking while anchoring the facts in the real world, blocks the *de dicto* reading which should actually become available when thinking slow. If a *de re* bias does exist, it would be plausible to consider that, provided *de dicto* is more accessible (as I suggested in section 2.1), when thinking fast *de dicto* is readily available and lawyers who think slow get to the *de re* reading thereby somehow blocking the previously accessible *de dicto*.

While Anderson admits that lawyers are not always blind to *de dicto* interpretations, she claims that lawyers are *de re* biased based on the four cases she presents. Further evidence for this *de re* bias comes from Rodes (1998). He presents 12 cases – some real and some made up – in which the *de re*/*de dicto* ambiguity plays a role and describes what courts ruled and how they often rule in similar court cases. In the majority of the cases the courts opt for a *de re* interpretation. 8 out of the 12 cases are real court cases. In 5 of the cases a *de re* ruling was chosen by the court, in 2 *de dicto* and in the one described in the introduction *de re* was opted for at the first instance but then overturned on a *de dicto* reading by the court of appeal. From a legal perspective, Rodes (1998:634) agrees with some of the rulings but describes others as "[unfortunate] and [...] quite [wrong]". Bix (2014) criticizes Anderson's paper because she only presents four court cases. Furthermore, these four cases are from different countries (and therefore different legal contexts) and while she does present a case from the 21st century, she also includes the fraud by impersonation case which occured in the 19th century.

Anderson's proposal that lawyers are de re biased is intriguing, however, both the evidence base and the explanation for the proposed cognitive bias are unsatisfactory. For this reason, it is important to conduct further research into whether lawyers are in fact *de re* biased.

2.3 Research Questions and Hypotheses

Bix's criticism of the data set used in Anderson (2014) prompts the main research question: Are lawyers *de re* biased? More specifically, are lawyers more likely to reject *de dicto* than non-lawyers? Furthermore, are lawyers more likely to accept *de re* than non-lawyers? These questions lead to the following two hypotheses based on the previous literature: Lawyers are more likely to reject *de dicto* than non-lawyers (who are not expected to reject *de dicto* at all). Lawyers are more likely to accept *de re* than non-lawyers – plausibly lawyers fall into the 45% of respondents who accepted *de re* in all cases in the study conducted by Zhang and Davidson. It is expected that a *de re* bias involves accepting both *de re* trials and rejecting both or at least one *de dicto* trial.

The second research question is whether there is a more accessible reading when the *de re*/*de dicto* ambiguity arises. More specifically, is the response pattern found in Zhang and Davidson (2021, 2024) replicable in the non-lawyer group of this dissertation? Additionally, is the response duration for one of the readings longer? In other words, if one reading is less accessible, it should take longer to process. So, is it the case that one reading takes more time? Zhang and Davidson's results suggest that *de dicto* is accessed more easily. Replicating these results and showing that response duration for *de dicto* is shorter than for *de re* would be further evidence for this. Therefore, it is hypothesized that response duration for *de dicto* will be shorter than for *de re* responses and that the pattern that *de dicto* is consistently accepted while *de re* has a bimodal distribution will be replicated.

3 Methodology

I replicated Zhang and Davidson's (2021) truth-value judgment experiment using the same materials but recruiting and comparing a different participant pool: lawyers and non-lawyers. In addition, I recorded the participants' response duration.

3.1 Participants

For this study, 50 participants were recruited who identified themselves as native English speakers and were at least 18 years old. Participants were recruited via targeted emails sent to UK based universities, law schools and law firms. Participants were not asked to specify what variety of English they spoke, but recruitment specifically targeted people living in the UK. The average participant was 33 years old (min = 19, max = 67). Additional languages spoken were French (14), German (13), Spanish (7), Japanese (3), and one each for Cantonese, Icelandic, Irish, Italian, Malay, Mandarin, Portuguese, Russian, and Turkish. No participant reported being diagnosed with a reading disability. Three participants reported an autism diagnosis. The participants were offered the opportunity to join a prize draw for a £30 Waterstones gift card upon completing the study to increase the number of participants and prevent dropouts.

Of the 50 participants, 20 had at least two years of formal legal education and were therefore considered lawyers by training/profession. This means that law students matching this condition were included in the study alongside lawyers. They studied/and practiced law between 2 (min) to 32 (max) years (mean: 9.1 years). Their specialization was commercial law (5), international law (3), corporate law (2), financial law(2), intellectual property (2), and one each in criminal law, family law, human rights, and medical law.

The control group of 30 participants (including all three reporting an autism diagnosis) reported educational backgrounds in: economics (4), business (3), physics (3), two each in architecture, history, languages, linguistics, philosophy, politics, and one each in film, humanities, journalism, mathematics, nanoscience, and psychology. One participant was a think piece and travel writer with no tertiary education.

3.2 Materials

All the material used was taken from Zhang and Davidson (2021). This means that there were four stories with four trials for each participant. As shown in (5-a) above, the stories were written in such a way that both a *de re* and a *de dicto* statement are theoretically true. Another example is given in (6-a) below (The full material used can be found in appendix I or at https://osf.io/qgnr5/ under "ZD.2020.appendix_1.exp1_material.pdf"). The context features a misapprehension where Mrs. Johnson believes the gift is for Annie when in fact it is for Grace. Therefore, the *de re* statement calls it *Grace's gift*, as can be seen in (6-b), whereas (6-c), the *de dicto* statement, calls it *Annie's gift*. As described above, the *de re* statement is *de re* in the sense that it is true on its *de re* interpretation. The same applies to the *de dicto* statement. There were three fillers per story of which one was definitely true (6-d), one was definitely false (6-e), and one was uncertain (6-f) based on the story.

(6) a. CONTEXT: Mr. and Mrs. Johnson have two high school girls, Annie and Grace. One day, Mrs. Johnson finds a wrapped present lying on the front porch of their house. A note on the box says: "From your secret admirer". Mrs. Johnson remembers that one day she saw Annie's classmate Mike standing in front of their house for a long time without knocking at the door. She also remembers being told that Annie is very popular in her class, so she concludes that Mike sent the gift to Annie. It turns out that Mike did send the gift, but to Grace. Grace and Mike met each other in a book club, and Mike has admired Grace since then.

b. Mrs. Johnson believes that Grace's gift was sent by Mike — DE RE

c. Mrs. Johnson believes that Annie's gift was sent by Mike — DE DICTO

d. When Mrs. Johnson finds the present, it is lying on the front porch with a note on it — DEFINITELY TRUE

e. Grace and Mike knew each other from jazz band — DEFINITELY FALSE

f. The gift was wrapped in pink paper — UNCERTAIN

(Zhang and Davidson 2021:Appendix 1)

As a novelty, the material was gamified to ensure that participants would not drop out and complete the experiment. The elements included a progress bar, feedback, and a score. The progress bar was chosen because it is easy to implement and has been shown to increase motivation (Mazarakis and Bräuer 2023). The feedback showed a green check mark for correct responses and a red cross for false responses. For the critical trials (the *de re* and *de dicto* trials), any response was rewarded with positive feedback. Feedback was chosen because it has been shown to increase motivation, especially when used in combination with other elements like a progress bar or scoring (Erhel and Jamet 2013, Mazarakis and Bräuer 2017 among others). Lastly, a running score showed participants how many answers out of the total 16 they had answered correctly. The final score was displayed upon submission of all responses.

3.3 Design & Procedure

First, all participants completed a brief biographical questionnaire (see appendix II for the complete questionnaire). Unlike in Zhang and Davidson (2024), each participant filled out the questionnaire before starting the experiment and not after. This was done to ensure that the questionnaire was filled out. Upon completion of the questionnaire, the participants were shown instructions followed by the first story. Participants got the chance to read each story completely before moving to the first judgment task, so the response duration could be measured accurately. The measurement of response duration started as soon as the participants saw the critical trial and ended when they submitted their judgment of that sentence. They were presented with the three fillers and one critical trial for each story. Each participant judged two *de re* and two *de dicto* statements. To get every possible order with two *de re* and two *de dicto* trials, following Zhang and Davidson, six lists were created achieving a Latin square design.[4] Participants were randomly assigned to one of the lists. Like in Zhang and Davidson (2024), the order of the stories and the order of the four trials within each story was randomized within each list. As in the Zhang and Davidson (2024) replication, participants used a Likert scale with five points to judge each sentence. The five points were: "Highly agree", "Somewhat agree", "Uncertain", "Somewhat disagree", and "Highly disagree". While there was no time pressure, participants could not return to previous stories or trials to change their answers, differing from Zhang and Davidson (2024). This was necessary for the response duration measurement and enabled access to the speakers' intuitions by getting their intuitive first response rather than a changed opinion at a later stage.

[4]The lists were (i) *de dicto*, *de dicto*, *de re*, *de re*; (ii) *de dicto*, *de re*, *de dicto*, *de re*; (iii) *de dicto*, *de re*, *de re*, *de dicto*; (iv) *de re*, *de dicto*, *de re*, *dicto*; (v) *de re*, *de re*, *de dicto*, *de dicto*; (vi) *de re*, *de dicto*, *de dicto*, *de re*.

4 Results

Only participants who answered at least 75% of the fillers correctly were included in the study. This should have lead to the exclusion of eleven participants. However, in story (5-a), repeated below in (7-a), the filler (7-b), which Zhang and Davidson (2021) considered to be definitely true, was reasonably answered differently by the majority of the participants. (7-a) calls Elizabeth a *younger and lesser-known* poet, which does not entail that Elizabeth is a *young* poet. Therefore, any response was counted as correct. No participant disagreed with (7-b). 23 participants responded "Uncertain", 11 "Somewhat agree", and 16 "Highly agree". Following this adjustment, seven of the eleven previously excluded participants reached the 75% threshold. Therefore, they were included and only four participants were excluded from the analysis. Out of the 46 included participants 18 were lawyers and 28 were non-lawyers.

(7) a. 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.

b. Elizabeth is a young poet

(Zhang and Davidson 2021:314)

4.1 Acceptance of *De Re* and *De Dicto*

The left half of Figure1 shows that in the *de dicto* condition the non-lawyers' judgment is "Highly agree" in most cases. For the *de re* condition nearly 25% of the non-lawyers' judgments were "Highly disagree" while the the largest share still goes to "Highly agree". These results support the results obtained by Zhang and Davidson (2021, 2024).

The right half of Figure1 shows that – with minor differences – lawyers show the same judgment pattern as non-lawyers; robust agreement in the *de dicto* condition and a bimodal distribution in the *de re* condition.

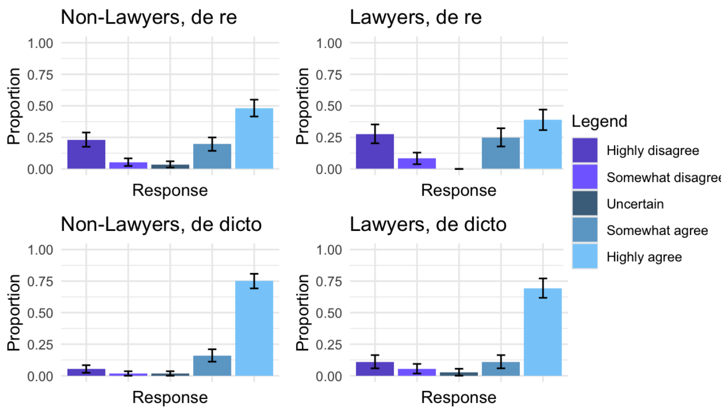

Figure 1: Proportion of responses by type and group

Following Zhang and Davidson (2024), the judgment data was fitted into a Bayesian multilevel cumulative ordinal model using the *brms* package (Bürkner et al. 2024) in R (see appendix III for the complete code and results of this study). The judgment levels on the Likert scale were recoded from 1 to 5 ("Highly disagree" = 1, "Highly agree" = 5) and were the dependent variable with non-equidistant intervals between them.[5] *De re* and *de dicto*, as the critical conditions, were entered as a dummy-coded fixed effect (*de dicto* was the reference level). Since this study included two groups of participants (lawyers and non-lawyers), the data was split in two. Furthermore, the subjects as well as the context were included as random effects. This way it became clear whether lawyers and non-lawyers differed in the two conditions, respectively. In the overall data the group as well as the interaction between group and critical condition was included as a fixed effect to determine differences between lawyers and non-lawyers.

Again following Zhang and Davidson, the prior distributions for all intercepts and fixed effect coefficients were set to follow a normal distribution with a mean of 0 and a standard deviation of 2 (i.e., Normal(0,2)). The prior for the correlation matrices was specified as LKJ(2), which is the default weakly informative prior for correlation matrices in the *brms* package

[5] The intervals are treated as non-equidistant because, for instance, it is not clear whether the distance between "Highly agree" and "Somewhat agree" is the same as the distance between "Somewhat agree" and "Uncertain".

(Lewandowski et al. 2009, Nalborczyk et al. 2019). The variances for the correlation matrices were kept at their default values in R. These priors imposed mild constraints on the possible coefficients for each parameter, while still permitting a reasonably large variance. The model was run with four sampling chains, each consisting of 4000 iterations, with the first 2000 iterations used for warmup. An $\hat{R}$ value close to 1.0 indicates that the sampling chains have successfully converged to the posterior distribution of the target predictor (Gelman and Rubin 1992). This setup follows previous acceptability ranking tasks from psycholinguistics (Paape et al. 2020, Zhang et al. 2023 among others).

All $\hat{R}$ values in the data of this study were 1.00 or 1.01 which marked a successful convergence. As in Zhang and Davidson (2024) the *emmeans* package (Lenth et al. 2024) was used to evaluate the main effects of the de re/de dicto manipulation and the judgment distinction across the groups. β is used to refer to the coefficient estimate and the highest posterior density (HPD) to the shortest interval with the highest density in the posterior distribution of the target coefficient (Box and Tiao 2011).

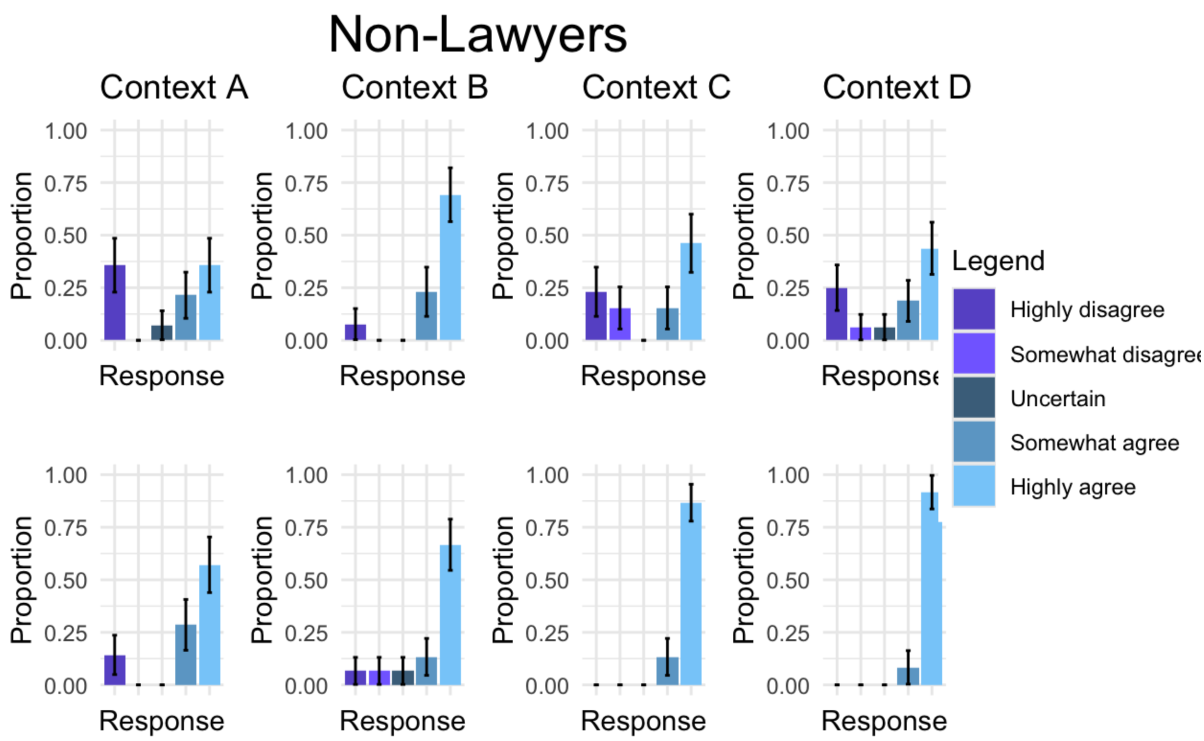

Figure 2: Judgment distribution across the four contexts (top: *de re*, bottom: *de dicto*)

Among non-lawyers *de dicto* was agreed with more than *de re* (β = 0.81, HPD = [0.357, 1.29]).[6] This replicates the findings presented in Zhang and Davidson (2024), although the difference is smaller (cf. β = 1.66, HPD = [0.16, 3.12] in Zhang and Davidson 2024:18). There was variation in the agreement distribution for *de re* and *de dicto* across contexts. Like in Zhang and Davidson (2024), there was no difference in context A (β = 1.32, HPD = [-1.41, 4.15]) or B (β = -0.324, HPD = [-2.89, 2.1]). While Zhang and Davidson found a difference in context C, no such difference was found among the non-lawyers (β = 2.25, HPD [-0.0481, 4.64]). Like in Zhang and Davidson's study, there was a difference between agreement with *de re* and *de dicto* in context D (β = 2.54, HPD [0.205, 5.1]). Although no difference could be found in context C in the Bayesian model, descriptively the judgment distribution for non-lawyers (see Figure2) resembles Zhang and Davidson's (2024) judgment distribution (see Figure**??**) very closely with contexts C and D receiving very high agreement on the *de dicto* condition while contexts A and B's agreement proportion is lower with "Highly agree" at around 60%. As in Zhang and Davidson's study, this experiment found the highest agreement for *de re* with "Highly agree" reaching nearly 75% in context B compared to below 50% in the other three contexts.

Among lawyers the *de dicto* condition also got higher agreement than the *de re* with the difference being slightly larger than among non-lawyers (β = 0.862, HPD = [0.261, 1.47]). In the overall data there was no difference between lawyers and non-lawyers (β = 0.229, HPD = [-0.173, 0.635]). For lawyers and non-lawyers, the judgment distribution varied by context. As for non-lawyers there was no difference between the agreement with the two conditions in contexts A (β = 0.601, HPD = [-2.51, 3.68]) and B (β = 1.47, HPD = [-4.06, 1]). In contrast to non-lawyers, lawyers showed a difference between *de re* and *de dicto* in context C (β = 3.53, HPD = [1.31, 5.94]), not however, in context D (β = 2.31, HPD = [-0.106, 4.97]). Figure 3 shows that, as with the non-lawyers and in Zhang and Davidson (2024), context B received higher agreement in the *de re* condition with "Highly agree" reaching ca 90% compared to below 50% in the other contexts. Moreover, like among non-lawyers, contexts C and D show a

[6] More difference is signified by higher a β-value. If the HPD crosses 0, there is no difference – if it does not, there is.

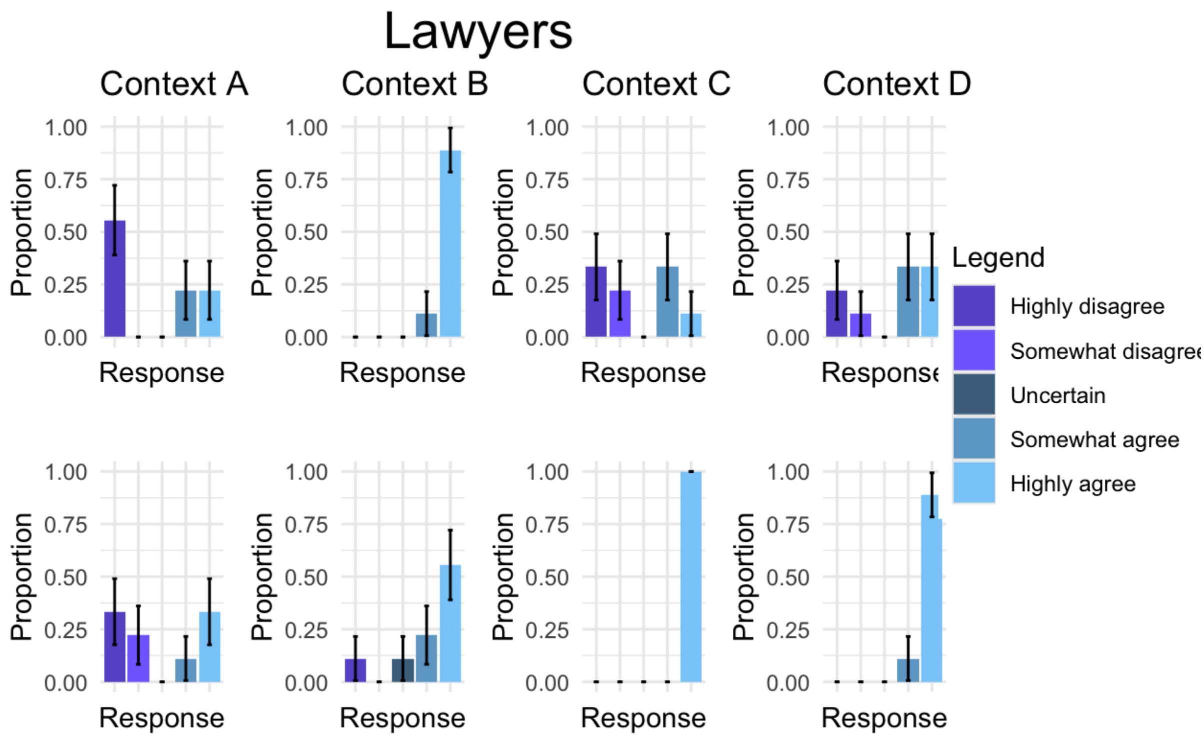

Figure 3: Judgment distribution across the four contexts (top: *de re*, bottom: *de dicto*)

very high agreement proportion in the *de dicto* condition with "Highly agree" reaching ca 90 – 100% compared to clearly below (A) and slightly above (B) 50%.

Additionally, the judgment data was fitted into a ordinal regression model using the *ordinal* package (Christensen 2023) in R. The data was separated by group again. The response was the dependent variable and the type (*de re* or *de dicto*) was the indenpendent variable. Subjects and contexts were included as random effects. Among non-lawyers there was a significant difference between the agreement with *de re* and *de dicto* with the *de re* condition receiving less agreement (SE = 0.4003, z = -3.166, p = 0.00155). Among lawyers there was also a significant difference between the agreement with the two types, again *de re* being agreed with less (SE = 0.5193, z = -2.609, p = 0.00907). Splitting the data by type (using response as the dependent variable, group as the independent variable, and subject and context as random effects) showed that there was no significant difference between lawyers and non-lawyers in the *de re* (SE = 0.4646, z = -0.744, p = 0.457) or the *de dicto* condition (SE = 0.6641, z = -0.895, p = 0.371).

	Agree with 0 trials	Agree with 1 trial	Agree with 2 trials	Total
de re	4 (14.3%)	10 (35.7%)	14 (50%)	28
de dicto	0 (0%)	5 (17.9%)	23 (82.1%)	28

Table 1: Distribution of non-lawyers by judgment behavior

	Agree with 0 trials	Agree with 1 trial	Agree with 2 trials	Total
de re	3 (16.7%)	7 (38.9%)	8 (44.4%)	18
de dicto	1 (5.6%)	5 (27.8%)	12 (66.7%)	18

Table 2: Distribution of lawyers by judgment behavior

Tables 1 and 2 show whether participants who agreed with one trial of either type also agreed with the second trial. In other words, did someone who accepted one *de re* trial also accept the other *de re* trial and did someone who accepted one *de dicto* trial also accept the other *de dicto*? Table 1 shows that 50% of the non-lawyers agreed with both *de re* statements while 82.1% agreed with both *de dicto* trials. While all non-lawyers agreed with at least one *de dicto* trial, Table 2 shows that one lawyer did not agree with either *de dicto* statement. Nevertheless, the judgment behavior for *de dicto* among the lawyers overall is similar to the non-lawyers. In the *de re* condition the lawyer's distribution also resembles the non-lawyer closely with around half agreeing with both and around a sixth agreeing with neither *de re* trial.

Only one participant rejected both *de dicto* trials and accepted both *de re* trials. However, six participants agreed with only one *de dicto* trial and both *de re* trials. Assuming that a *de re* bias involves agreeing with both *de re* trials and not agreeing with both *de dicto* statements, there are thus seven participants (three non-lawyers and four lawyers) who are *de re* biased in this study. Out of the five non-lawyers who agreed with only one *de dicto* trial (see Table 1), three agreed with both *de re* trials. These three non-lawyers' educational background is languages, economics, and architecture, respectively. Exactly as among the non-lawyers, out of the five lawyers who agreed with only one *de dicto* trial, three agreed with both *de re* trials. Furthermore, the lawyer who agreed with neither *de dicto* (see Table 2) trial agreed with both *de re*. The four lawyers who agreed with both *de re* trials and only one or neither of the *de dicto* trials specialize in international law, criminal and family law, litigation of commercial disputes, and corporate law, respectively and have 3, 5, 7, and 12 years of legal experience. Altogether, the seven participants who agreed with both *de re* and one or neither *de dicto* trial are 23 (2), 22, 26, 44, 49, and 57 years old. They are monolinguals (2), bilinguals (French: 1, Icelandic: 1), a trilingual (French and Spanish), a quadrilingual (French, German, and Russian), and a speaker of five languages (French, German, Spanish and Japanese). To conclude, the qualitative analysis of the participants with a *de re* bias shows no connection based on any of the parameters collected in the questionnaire.

4.2 Response Duration

As is standard in the literature, an a priori screening of outliers that are at least two standard deviations from the mean were excluded from the analysis (Baayen and Petar 2010).[7] Seven observations were excluded this way. 177 observations were included in the analysis.

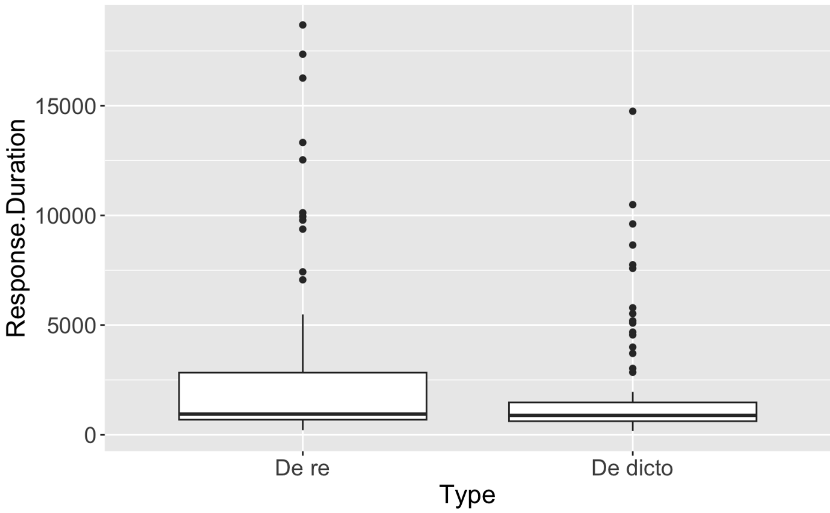

Figure 4: Response duration for *de re* and *de dicto* (overall)

Figure 4 shows that *de dicto* response durations were quicker than for *de re* across the 46 participants. The mean for *de re* was 2791.8ms compared to 1859.0ms for *de dicto*. Overall, the shortest response duration was also a *de dicto* trial at 174.0ms compared to the shortest *de re* response duration at 210.0ms. The longest overall response duration was a *de re* trial with 18684.2ms compared to the slowest *de dicto* response duration at 14749.0ms. The data was fitted into a linear mixed effects model using the *lmerTest* package (Kuznetsova et al. 2022). Response duration was the dependent variable, type the independent variable and participant and context were random effects. It could be shown that the response duration for *de dicto* was

[7]Outliers in response duration were not excluded from the judgment data analysis, since it is plausible that participants got distracted and took longer with no relevant influence on the judgment.

significantly shorter than for *de re* (SE = 461.24, t = -2.039, p = 0.0436). The response had no significant effect on the response duration. In other words, "Uncertain" did not lead to a significantly shorter or longer response duration compared to the four other response options. This underlines that *de dicto* response durations are truly shorter than *de re* independent of the fact that *de dicto* was more widely accepted than *de re*, as was shown in section 4.2.

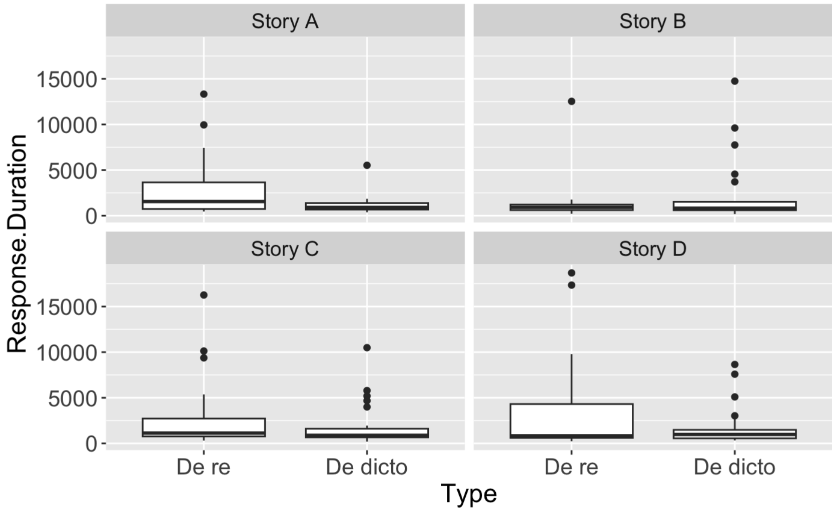

Figure 5: Response duration for *de re* and *de dicto* by context

There was some difference between response durations in the four different contexts, as can be seen in Figure5. The average *de dicto* response duration in context A was 1184.4ms contrasted with 2959.7ms for *de re*. *De dicto* response duration was only significantly shorter than *de re* response duration in context A (SE = 730.768, t = -2.446, p = 0.0187). In the other three scenarios there was no significant difference between the response duration of *de re* and *de dicto* (B: SE = 954.7, t = 0.95, p = 0.3475; C: SE = 1014.1, t = -1.057, p = 0.296745; D: SE = 1235.9, t = -1.378, p = 0.175518). Notably, the average *de dicto* response duration was only shorter than the average *de re* response duration in contexts A, C, and D. By contrast, the mean *de re* response duration (1470.3ms) was shorter than the average *de dicto* response duration (2377.4ms) in context B.

Lastly, Figure 6 shows that the average response duration for *de dicto* is shorter than for *de re* in both groups. The mean response duration for *de dicto* among lawyers is 1172.0ms compared to 2701.3ms for *de re*. Among lawyers the response duration for *de dicto* was significantly shorter than for *de re* (SE = 686.37, t = -2.228, p = 0.0302). There was no significant difference

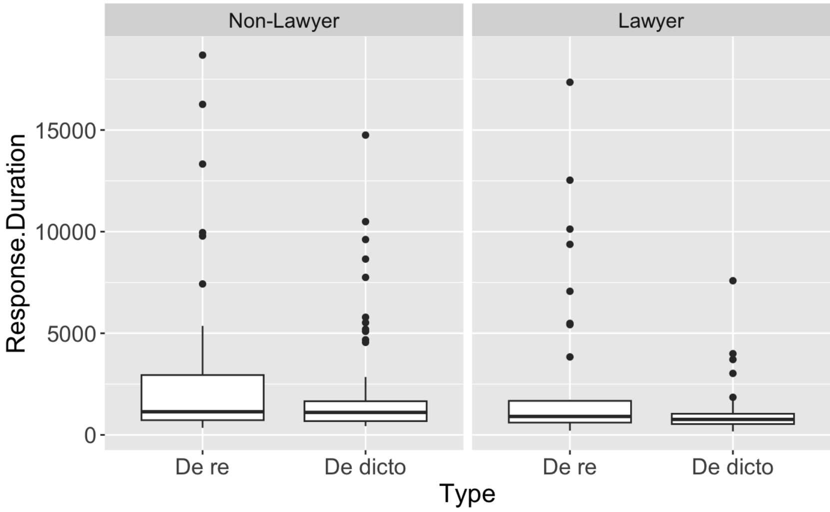

Figure 6: Response duration for *de re* and *de dicto* by group

between the response duration for *de re* and *de dicto* among non-lawyers (SE = 612.18, t = -0.901, p = 0.371). There was no significant difference between lawyers and non-lawyers for response duration overall (SE = 618.61, t = -1.174, p = 0.248). There was also no significant difference in response duration between lawyers and non-lawyers for the *de re* condition (SE = 997.15, t = -0.265, p = 0.793). However, for the *de dicto* condition, lawyers were significantly faster than non-lawyers (SE = 536.3, t = -2.135 p = 0.0355).

5 Discussion

5.1 Are *De Re* and *De Dicto* equally accessible?

Since the difference between *de re* and *de dicto* responses observed by Zhang and Davidson (2024) was replicated among non-lawyers (and lawyers) and it was shown that *de dicto* is accepted significantly more by participants than *de re*, it seems like *de dicto* might be more accessible. This is especially striking since the population in this study came from a different language community than Zhang and Davidson's samples. Therefore, the hypothesis that *de dicto* is consistently accepted while *de re* shows a bimodal distribution and Zhang and Davidson's results were replicated would be confirmed. Furthermore, the fact that more than 80% of the non-lawyers agreed with both *de dicto* trials and all with at least one, while only 50% agreed with both *de re* trials, points towards *de dicto* being accessed more easily.

However, closer examination of the differences – or the lack thereof – across the four contexts calls this conclusion into question. If *de dicto* were truly more accessible than *de re*, this should be observable across all scenarios, however, only context D showed this difference among non-lawyers. Moreover, even though the response duration in the *de dicto* condition is shorter overall – which would be an indication of *de dicto* being more accessible – these findings were not significant. Again, a closer examination of the contexts shows that only in context A the *de dicto* condition showed a significantly shorter response duration. Again, this should have been the case across all contexts if *de dicto* was truly more accessible. Furthermore, in the comparison by group, it became clear that the *de dicto* response duration was actually only significantly shorter among lawyers. Seeing as there was no significant difference between the two conditions among non-lawyers, there seems to be little evidence for the fact that *de dicto* is more accessible than *de re*.

The judgment disparity between von Fintel and Heim and Nelson introduced above can be explained by the fact that while most participants agreed with both *de re* trials, there was a large portion of non-lawyers (35.7%) who only agreed with one of the trials and some (14.3%) who agreed with neither. As Zhang and Davidson (2021) suggest, it is plausible that a majority of the population agrees with *de re* consistently, while some people do not.

In sum, the results of this study do not suggest that either *de dicto* or *de re* is more acceptable than the other. This is in line with the different approaches in the formal semantics literature which do not make any claims that one would be more accessible than the other (Zhang and Davidson 2024). Still, in light of the differences found across contexts in both judgment distribution and response duration suggest that the models need to take into account the intricacies and subtle differences between contexts. Further quantitative research will be necessary to uncover the exact judgment distributions of the *de re/de dicto* ambiguity.

5.2 Are Lawyers Like Everyone Else?

No indication of lawyers being *de re* biased was found in the present study, since lawyers' judgment patterns did not differ from non-lawyers' judgment distribution. The hypothesis that *de dicto* would receive less agreement among lawyers compared to non-lawyers was disproved as there are no significant differences between the two groups in agreeing in the *de dicto* condition. Although the only participant who agreed with neither *de dicto* trial was a lawyer, this is not sufficient evidence to support the hypothesis of a *de re* bias among lawyers. It was also shown that – contrary to the hypothesis – lawyers did not agree with *de re* trials more than non-lawyers. To the contrary, the response duration data suggests that lawyers are actually not *de re* biased at all. Whereas there was non significant difference in response duration for non-lawyers, among lawyers response duration was significantly shorter for *de dicto* compared to the *de re* condition. This indicates that lawyers access *de dicto* more easily than *de re* which goes against the idea of the *de re* bias. Furthermore, in the *de dicto* condition lawyers were significantly quicker than non-lawyers, which hints at the fact that lawyers access *de dicto* more easily than non-lawyers.

Although the results of this study do not suggest that lawyers are *de re* biased, it should be mentioned that it is commonly accepted that register plays a role in language use (Wiese et al. 2022, Pescuma et al. 2023, Rotter and Liu 2024 among others). Pescuma et al. (2023) define registers as the different language patterns which are used in a recurrent and conventionalized way by a speech community dependent on the situational-functional context. So even though the results in this study give no indication of a *de re* bias among lawyers, it is possible that – if such a bias exists – it only surfaces among lawyers when they are in legal contexts, for example, in court rooms or when preparing for a court case. The present study did not feature register differences, so further research is necessary to confirm whether a *de re* bias exists. Since Bix criticizes Anderson's paper for only considering four court cases from different countries and centuries, it is not surprising that this study found no indication of a *de re* bias. In other words, it is plausible that it is not lawyers who show a *de re* bias but merely Anderson's sample of court cases. This would support the conclusion that lawyers' judgment patterns with respect to the *de re*/*de dicto* ambiguity do not differ from everyone else.

5.3 Individual and Context-Dependent Differences

As in Zhang and Davidson (2021, 2024), it was shown that there are differences between speakers. While most participants agreed with both *de dicto* trials, some agreed with just one and one lawyer even agreed with neither. Additionally, while most speakers agreed with both statements in the *de re* condition, some only agreed with one or neither. Taking a more qualitative look into the *de re* bias and the idea that it involves a rejection of *de dicto* and agreement with *de re*, showed that the participants who did not agree with both trials in the *de dicto* condition were more likely to agree with both *de re* trials. According to this definition four lawyers and three non-lawyers classified as *de re* biased. However, neither their legal specialization – they all specialized in different fields – nor their years of experience in law – spanning 3–12 years – proved to be an indicator of these responses. In other words, a specific specialization or a specific amount of experience in law did not lead participants to reject *de dicto* and agree with *de re*. The background information collected could also not deliver any clues in explaining who may be *de re* biased among non-lawyers. There are three other economics, two other languages, and another architecture student/graduate who all agreed with both trials in the *de dicto* condition. Furthermore, although not enough data was gathered to confirm this, it does not seem that the *de re* bias is generational, since the seven participants were between 22 and 57 years old and as such not part of the same generation. There was also no commonality with regard to the languages spoken by the seven *de re* biased participants as they included mono-, bi-, and multilinguals who spoke a variety of different languages. In essence, although there are obvious differences between speakers and some exhibit what could be called a *de re* bias, neither legal training nor any of the other parameters gathered in the questionnaire proved indicative of this *de re* bias. It is possible that regional variety has an influence, however, this seems unlikely as the judgment distribution in this study was similar to the one observed by Zhang and Davidson which was conducted in the USA. It would be fairly surprising if the variety of English in the sample of this UK-based study was so similar to Zhang and Davidson's.

Differences in the judgment distribution across contexts suggest that context plays a vital role in the availability of *de re* and *de dicto*. This is further supported by the differences in response duration across the different contexts in the two conditions – even though these differences were only significant in context A. The fact that response duration in the *de dicto* condition was significantly shorter than in the *de re* condition in context A could suggest that this context eases the agreement with *de dicto*. However, no significant difference was found in terms of the agreement with *de re* or *de dicto* in this context in the Bayesian model, replicating the results of Zhang and Davidson (2024).

In the *de re* condition context B received unusually high agreement with the proportions of "Highly" and "Somewhat disagree" being the lowest compared to the other contexts. This replicates findings from both studies by Zhang and Davidson. In their 2024 paper, they suggest that this may be explained by the information structure of the *de re* trial in context B. That is, *Mrs Johnson believes that Grace's gift was sent by Mike* includes a passive construction (recall (6-a) for the full context). More specifically, they point out that the passive, and particularly, the *by* phrase stresses that it was Mike's gift (for Grace) and that the verb *sent* further stresses that it was by Mike for Grace, contrasting with *received*, for instance. However, they note that context D also features a passive, i.e. *Tracy believes that Alice's spare apron needs to be washed* and that context D receives higher disagreement in the *de re* condition than B – a finding that is replicated in this study. They put forward the idea that the *by* phrase is the deciding factor which leads to the high agreement in context B, but leave this question open to be discussed by further research. Since the judgment distribution observed by Zhang and Davidson (2024) was replicated in this study, there is strong evidence that context B/the *de re* trial in context B eases agreement with *de re*. Although the difference was not significant, the fact that the response duration for *de re* was shorter than for *de dicto* in context B could correlate with the high acceptance in the *de re* condition in this context. However this remains unclear as overall the response had no significant effect on response duration. The *by* phrase explanation that context B received higher agreement in the *de re* condition than the other contexts seems plausible, nevertheless, this question will need to be explored further in future research.

It is easily explained that there was no significant difference in context B in the Bayesian model among the lawyers and non-lawyers, since the agreement with *de re* was very high in this condition. While Zhang and Davidson (2024) found differences between the two conditions in contexts C and D, this study only found a difference in context D among non-lawyers and a difference in context C among lawyers. However, in both groups of this study, agreement with *de dicto* is noticeably higher in contexts C and D compared to contexts A and B. Zhang and Davidson give no insight as to why they only find a difference between the conditions in these two contexts. Context C differs from the other contexts in that it is about humans, i.e. Haley's brother/husband, whereas the other contexts are about things, i.e. a poem, a gift, and an apron, respectively. This cannot be the explanation, since context D is like contexts A and B.

(8) a. Julie believes that [Nicole's]$_{\text{de dicto}}$/[Elizabeth's]$_{\text{de re}}$ poem will win the competition.
b. Mrs. Johnson believes that [Annie's]$_{\text{de dicto}}$/[Grace's]$_{\text{de re}}$ gift was sent by Mike.
c. Susan believes that Haley's [brother]$_{\text{de dicto}}$/[husband]$_{\text{de re}}$ will accompany her for a while.
d. Tracy believes that Alice's [favorite]$_{\text{de dicto}}$/[spare]$_{\text{de re}}$ apron needs to be washed.

However, there is a striking difference between contexts A and B and contexts C and D. (8) shows the sentences in the *de dicto* and *de re* conditions for each of the four contexts. While in context A (8-a) and B (8-b) the difference between the *de re* and *de dicto* statement is the possessor, in contexts C (8-c) and D (8-d) it is about the possessee/possessed thing. In other words, while contexts A and B are about whether it is Nicole or Elizabeth who will win or Annie or Grace's gift, contexts C and D are always about Haley or Alice, respectively. The question in these two contexts is whether it is the husband or brother/spare or favorite apron. So, it is possible that it was not the context, but rather the specific statements which led to the observed judgment distribution. Further research, focusing on the possessor/possessee distinction, needs to be conducted to determine whether it truly influences the acceptance of *de re* and *de dicto*. The results of this research are important for the improvement of the currently available approaches to the *de re*/*de dicto* ambiguity in the formal semantics literature. Based on the findings in this study it is expected that it influences the acceptability of *de dicto*, not however of *de re*, since *de re* was highly accepted in context B and lower in the other three contexts, whereas *de dicto* showed high agreement in contexts C and D and lower agreement (although still more agreement than disagreement) in contexts A and B. In short, agreement with both *de re* and *de dicto* seems to be susceptible to seemingly minor differences in the information structure, with *de re* probably being influenced by the absence or presence of a *by* phrase and *de dicto* by the possessor/possessee distinction.

6 Conclusion

To sum up, in this study I replicated Zhang and Davidson's (2021, 2024) truth-value judgment task evaluating the agreement with *de re* and *de dicto* statements in contexts in which both are theoretically true. Additionally, I compared lawyers and non-lawyers and recorded the response duration for the two conditions to answer the following two questions. Firstly, is *de dicto* more accessible than *de re*? And secondly, are lawyers *de re* biased? I found no evidence to support that *de dicto* is more accessible than *de re* or that lawyers are *de re* biased. A *de re* bias could also not be explained by the educational background, legal specialization, or experience in law among the participants who agreed consistently with *de re* but not with *de dicto*. However, a difference between the *de re* statements, namely the *by* phrase, was found to potentially explain the differences in agreement proportion across contexts in the *de re* condition. Similarly, among the *de dicto* trials, a difference in whether the possessor or possessee showed the contrast between *de re* and *de dicto* was found to likely influence the proportion of agreement with a trial. Rounding off, the results of this study suggest that – from a linguistic point of view – it should not be surprising that in Regina v. Smith (David) [1974], Smith was acquitted on a *de dicto* interpretation by the court of appeal.

References

Anderson, Jill C. (2014). Misreading Like a Lawyer: Cognitive Bias in Statutory Interpretation. *Harvard Law Review 127*(6), 1521–1592.

Apperly, Ian A., Elisa Back, Dana Samson, and Lisa France (2008). The cost of thinking about false beliefs: evidence from adults' performance on a non-inferential theory of mind task. *Cognition 106*(3), 1093–1108.

Baayen, Harald and Milin Petar (2010). Analyzing reaction times. *International Journal of Psychological Research 3*(2), 12–28.

Birch, Susan A.J. and Paul Bloom (2004). Understanding children's and adults' limitations in mental state reasoning. *Trends in Cognitive Sciences 8*(6), 255–260.

Birch, Susan A.J. and Paul Bloom (2007). The Curse of Knowledge in Reasoning About False Beliefs. *Psychological Science 18*(5), 382–386.

Bix, Brian H. (2014). Statutory Reading of Opaque Constructions — Errors and Purposes. *Harvard Law Review 127*(6), 200–203.

Box, George E. P. and George C. Tiao (2011). *Bayesian Inference in Statistical Analysis*. Wiley.

Bürkner, Paul-Christian, Jonah Gabry, Sebastian Weber, Andrew Johnson, Martin Modrak, Hamada S. Badr, Frank Weber, Aki Vehtari, Mattan S. Ben-Shachar, Hayden Rabel, Simon C. Mills, Stephen Wild, and Ven Popov (2024). Bayesian Regression Models using 'Stan'. *R package version 2.21.0*.

Capone, Alessandro (2011). Default semantics and the architecture of the mind. *Journal of Pragmatics 43*(6), 1741–1754.

Christensen, Rune Haubo Bojesen (2023). Regression Models for Ordinal Data. *R package version 2023.12.4*.

Cresswell, Maxwell J. and Arnim von Stechow (1982). *De re* belief generalized. *Linguistics and Philosophy 5*, 503–535.

Criminal Damage Act (1971). c. 48, s. 1. In *UK Public General Acts*. legislation.gov.uk. URL https://www.legislation.gov.uk/ukpga/1971/48/section/1, Access: 2024-07-28.

Erhel, S. and E. Jamet (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers Education 67*, 156–167.

Fodor, Janet D. (1970). *The Linguistc Description of Opaque Contexts*. Ph. D. thesis, MIT.

Gelman, Andrew and Donald B. Rubin (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science 7*, 457–472.

ICLR (2024). Mens rea and actus reus. In *Glossary*. The Incorporated Council of Law Reporting for England and Wales (ICLR). URL https://www.iclr.co.uk/knowledge/glossary/mens-rea-and-actus-reus/, Access: 2024-07-31.

Jaszczolt, Kasia M. (1997). The 'default de re' principle for the interpretation of belief utterances. *Journal of Pragmatics 28*(3), 315–336.

Jaszczolt, Kasia M. (2015, 02). Default Semantics. In *The Oxford Handbook of Linguistic Analysis*. Oxford University Press.

Kahneman, Daniel (2012). *Thinking, Fast and Slow*. London: Penguin Books Ltd.

Kinderman, Peter, Robin Dunbar, and Richard P. Bentall (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology 89*(2), 191–204.

Kuznetsova, Alexandra, Per Bruun Brockhoff, Rune Haubo Bojesen Christensen, and Sofie Pødenphant Jensen (2022). Tests in Linear Mixed Effects Models. *R package version 3.1.3*.

Lenth, Russell V., Ben Bolker, Paul Buerkner, Iago Giné-Vázquez, Maxime Herve, Maarten Jung, Jonathon Love, Fernando Miguez, Julia Piaskowski, Hannes Riebl, and Henrik Singmann (2024). Estimated Marginal Means, aka Least-Squares Means. *R package version 1.10.3*.

Lewandowski, Daniel, Dorota Kurowicka, and Harry Joe (2009). Generating Random Correlation Matrices Based on Vines and Extended Onion Method. *Journal of Multivariate Analysis 100*(9), 1989–2001.

Mazarakis, Athanasios and Paula Bräuer (2017). Welche Gamification motiviert? In *Workshop Gemeinschaften in Neuen Medien (GeNeMe) 2017*, pp. 259–268. Dresden: TUDpress.

Mazarakis, Athanasios and Paula Bräuer (2023). Gamification is Working, but Which One Exactly? Results from an Experiment with Four Game Design Elements. *International Journal of Human–Computer Interaction 39*(3), 612–627.

Nalborczyk, Ladislas, Cedric Batailler, Helene Loevenbruck, Anne Vilain, and Paul-Christian Burkner (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.

Nelson, Michael (2019). The *de re/de dicto* distinction (Supplement to Propositional Attitude Reports). In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. URL https://plato.stanford.edu/entries/prop-attitude-reports/dere.html, Access: 2024-03-05.

Paape, Dario, Shravan Vasishth, and Titus von der Malsburg (2020). Quadruplex Negatio Invertit? The On-Line Processing of Depth Charge Sentences. *Journal of Semantics 37*(4), 509–555.

Pescuma, Valentina N., Dina Serova, Julia Lukassek, Antje Sauermann, Roland Schäfer, Aria Adli, Felix Bildhauer, Markus Egg, Kristina Hülk, Aine Ito, Stefanie Jannedy, Valia Kordoni, Milena Kuehnast, Silvia Kutscher, Robert Lange, Nico Lehmann, Mingya Liu, Beate Lütke, Katja Maquate, Christine Mooshammer, Vahid Mortezapour, Stefan Müller, Muriel Norde, Elizabeth Pankratz, Angela G. Patarroyo, Ana-Maria Pleşca, Camilo R. Ronderos, Stephanie Rotter, Uli Sauerland, Gohar Schnelle, Britta Schulte, Gediminas Schüppenhauer, Bianca Maria Sell, Stephanie Solt, Megumi Terada, Dimitra Tsiapou, Elisabeth Verhoeven, Melanie Weirich, Heike Wiese, Kathy Zaruba, Lars Erik Zeige, Anke Lüdeling, and Pia Knoeferle (2023). Situating language register across the ages, languages, modalities, and cultural aspects: Evidence from complementary methods. *Frontiers in Psychology 13*, 964658.

Regina v. Smith (David) (1974). Q.B. 354.

Rodes, Robert E. (1998). De re and de dicto. *Notre Dame Law Review 73*(3), 627–636.

Roeper, Tom (2007). *The Prism of Grammar: How Child Language Illuminates Humanism*. The MIT Press.

Romoli, Jacopo and Yasutada Sudo (2009). De Re/De Dicto Ambiguity and Presupposition Projection. In *Proceedings of Sinn und Bedeutung 13*, pp. 425–438.

Rotter, Stephanie and Mingya Liu (2024). A register approach to negative concord versus negative polarity items in english. *Linguistics*. URL https://doi.org/10.1515/ling-2023-0016, Access: 2024-08-15.

Russell, Bertrand (1905). On Denoting. *Mind 14*(56), 479–493.

Rutherford, M. D. (2004). The effect of social role on theory of mind reasoning. *British Journal of Psychology 95*(1), 91–103.

Shrout, Patrick E. and Joseph L. Rodgers (2018). Psychology, Science, and Knowledge Construction: Broadening Perspectives from the Replication Crisis. *Annual Review of Psychology 69*, 487–510.

Sönning, Lukas and Valentin Werner (2021). The Replication Crisis, Scientific Revolutions, and Linguistics. *Linguistics 59*(5), 1179–1206.

von Fintel, Kai and Irene Heim (2011). Intensional semantics. Lecture notes.

Whiteley v. Chappell (1868). L.R. 4 Q.B. 147.

Wiese, Heike, Artemis Alexiadou, Shanley Allen, Oliver Bunk, Natalia Gagarina, Kateryna Iefremenko, Maria Martynova, Tatiana Pashkova, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehaye, Sabine Zerbian, and Yulia Zuban (2022). Heritage Speakers as Part of the Native Language Continuum. *Frontiers in Psychology 12*, 717973.

Zhang, Yuhan and Kathryn Davidson (2021). De re interpretation in belief reports: An experimental investigation. In *Proceedings of ELM 1*, pp. 310–321.

Zhang, Yuhan and Kathryn Davidson (2024). Interpreting referential noun phrases in belief reports – the de re/de dicto competition. In press.

Zhang, Yuhan, Rachel Ryskin, and Edward Gibson (2023). A Noisy-Channel Approach to Depth-Charge Illusions. *Cognition 232*, 105346.

Appendix I

Experiment Material (Judgment Task)

Complete experiment material from experiment 1 in “De re interpretation in belief reports – An experimental investigation” (Zhang and Davidson 2021) as it was used in this study (instructions were changed from “slider bar” to “scale”:

Context A

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.

Instructions: According to this story, please use the scale to indicate to what extent you agree or disagree with the following four statements.

de dicto: Julie believes that Nicole’s poem will win the competition.

de re: Julie believes that Elizabeth’s poem will win the competition.

definitely right: Elizabeth is a young poet.

definitely wrong: Elizabeth and Nicole met each other and decided that they will both write poems about the ocean.

not sure: Julie will also be the judge for the poetry competition next year.

Context B

Context: Mr. and Mrs. Johnson have two high school girls, Annie and Grace. One day, Mrs. Johnson finds a wrapped present lying on the front porch of their house. A note on the box says: “From your secret admirer”. Mrs. Johnson remembers that one day she saw Annie’s classmate Mike standing in front of their house for a long time without knocking at the door. She also remembers being told that Annie is very popular in her class, so she concludes that Mike sent the gift to Annie. It turns out that Mike did send the gift, but to Grace. Grace and Mike met each other in a book club, and Mike has admired Grace since then.

Instructions: According to this story, please use the scale to indicate to what extent you agree or disagree with the following four statements.

de dicto: Mrs. Johnson believes that Annie’s gift was sent by Mike.

de re: Mrs. Johnson believes that Grace’s gift was sent by Mike.

definitely right: When Mrs. Johnson finds the present, it is lying on the front porch with a note on it.

definitely wrong: Grace and Mike knew each other from jazz band.

not sure: The gift was wrapped in pink paper.

Context C

Context: Susan works at a hospital. She is responsible for checking in visitors whose relatives and friends are in the maternity ward. One day, a man comes to Susan and asks to visit Haley. His surname is the same as Haley's and they both have beautiful blond hair. Susan remembers Haley saying that she has a brother, so Susan concludes that this man is Haley's brother. Since Haley will deliver a little baby soon, Susan also thinks that the man will accompany Haley for a while. Yet, it turns out that this man is not Haley's brother but instead, Haley's husband. Haley took her husband's surname, and they both have blond hair.

Instructions: According to this story, please use the scale to indicate to what extent you agree or disagree with the following four statements.

de dicto: Susan believes that Haley's brother will accompany her for a while.

de re: Susan believes that Haley's husband will accompany her for a while.

definitely right: Haley is receiving medical care in the maternity ward.

definitely wrong: Susan thinks the man is related to Haley because of his brown hair.

not sure: The man is bringing a bouquet of daisies to Haley.

Context D

Context: Alice and Tracy live in the same apartment and always help each other with daily errands. One day, Tracy is gathering up their laundry and she finds an apron with a large coffee stain lying on the sofa. Tracy remembers Alice saying that she usually wears her favourite apron when she cooks and the other day she spilled a cup of coffee while cooking. Tracy thus concludes that what she found is Alice's favourite apron and it needs to be washed. As a matter of fact, however, what Tracy found is Alice's spare apron, not her favourite one. Alice's favourite apron was already in the laundry at the time when she spilled the coffee onto her spare apron.

Instructions: According to this story, please use the scale to indicate to what extent you agree or disagree with the following four statements.

de dicto: Tracy believes that Alice's favourite apron needs to be washed.

de re: Tracy believes that Alice's spare apron needs to be washed.

definitely right: Alice usually wears an apron when she cooks.

definitely wrong: The apron with a large coffee stain was lying on the table when Tracy discovered it.

not sure: Tracy altogether gathered three pounds of laundry.

Appendix II

Questionnaire

Complete questionnaire used for this study:

- How old are you? (Please provide your answer in years)
- If you speak any language other than English, which one(s)?
- Have you ever been diagnosed with any reading disorders (e.g. dyslexia)?
- Have you ever been diagnosed with autism?
- Are you studying/Have you studied law? (yes)
  - Have you completed at least two years of university courses in law?
    - What area(s) of law do you specialise in?
- Are you studying/Have you studied law? (no)
  - Are you attending or have you attended university? (yes)
    - My degree(s) is/was in
      - Biology
      - Business
      - Chemistry
      - Computer Science
      - Economics
      - Finance
      - Linguistics
      - Languages
      - Mathematics
      - Medicine
      - Philosophy

      - · Sociology
      - · Other – please specify
  - Are you attending or have you attended university? (no)
    - ∗ What is your field of work?

Appendix III

Detailed Code and Results for Bayesian Model of Judgment Task

Are lawyers overall different from non-lawyers?

```
```{r}
all_data$Type = factor(all_data$Type, levels = c("De dicto", "De re"))
all_data <- all_data %>%
  mutate(Response = recode(Response, "Highly disagree" = 1, "Somewhat disagree" = 2, "Uncertain" = 3, "Somewhat agree" = 4, "Highly agree" = 5))
prior_rating <- c(brms::set_prior("normal(0,2)",class="b"),
                  brms::set_prior("normal(0,2)",class="Intercept"))
                  # set_prior("lkj(2)", class = "cor"))
m11 = brms::brm(Response ~Type*Group+(1|ParticipantID)+(1|Context),
         data=all_data,
         family = brms::cumulative(link ="probit",
         threshold="flexible"),
         init=0,
         prior=prior_rating,
         iter=4000,
         cores=2)
summary(m11)
```
Family: cumulative
  Links: mu = probit; disc = identity
Formula: Response ~ Type * Group + (1 | ParticipantID) + (1 | Context)
   Data: all_data (Number of observations: 184)
  Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
         total post-warmup draws = 8000
Multilevel Hyperparameters:
~Context (Number of levels: 4)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)   0.64   0.48   0.16   1.87 1.00 1259   721
```

```
~ParticipantID (Number of levels: 46)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     0.25      0.16     0.01     0.60 1.00     2134     3684
Regression Coefficients:
                     Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept[1]            -1.57      0.38    -2.28    -0.69 1.00     2240     1071
Intercept[2]            -1.35      0.38    -2.04    -0.48 1.00     2205     1095
Intercept[3]            -1.24      0.38    -1.93    -0.40 1.00     2181     1066
Intercept[4]            -0.65      0.38    -1.34     0.20 1.00     2153     1024
TypeDere                -0.82      0.24    -1.31    -0.35 1.00     4334     1941
GroupLawyer             -0.24      0.29    -0.81     0.34 1.00     5522     5779
TypeDere:GroupLawyer     0.02      0.37    -0.70     0.74 1.00     6086     5880
Further Distributional Parameters:
     Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
disc     1.00      0.00     1.00     1.00   NA       NA       NA
Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```{r}
em2 = emmeans::emmeans(m11, specs=pairwise~Group)
em2$contrasts %>%
  summary(infer=T)
```
 contrast   estimate lower.HPD upper.HPD
 (Non-Lawyer) - Lawyer   0.229   -0.173   0.635
Results are averaged over the levels of: Type
Note: contrasts are still on the probit scale
Point estimate displayed: median
HPD interval probability: 0.95
```

Difference between type among non-lawyers

```
```{r}
```
```

````
all_data_NonLawyers <- all_data[which(all_data$Group == "Non-Lawyer"),]
all_data_NonLawyers$Type = factor(all_data_NonLawyers$Type, levels = c("De dicto", "De re"))
prior_rating <- c(brms::set_prior("normal(0,2)",class="b"),
                 brms::set_prior("normal(0,2)",class="Intercept"))
                 # set_prior("lkj(2)", class = "cor"))
m12 = brms::brm(Response ~Type+(1|ParticipantID)+(1|Context),
         data=all_data_NonLawyers,
         family = brms::cumulative(link ="probit",
         threshold="flexible"),
         init=0,
         prior=prior_rating,
         iter=4000,
         cores=2)
summary(m12)
```
 Family: cumulative
  Links: mu = probit; disc = identity
Formula: Response ~ Type + (1 | ParticipantID) + (1 | Context)
   Data: all_data_NonLawyers (Number of observations: 112)
  Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
         total post-warmup draws = 8000
Multilevel Hyperparameters:
~Context (Number of levels: 4)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     0.45      0.46     0.02     1.91 1.02      225       47
~ParticipantID (Number of levels: 28)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     0.24      0.18     0.01     0.65 1.00     1698     2715
Regression Coefficients:
             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept[1]    -1.56      0.42    -2.22    -0.19 1.02      164       31
Intercept[2]    -1.36      0.40    -2.00    -0.03 1.02      166       31
Intercept[3]    -1.22      0.41    -1.86     0.19 1.02      165       31
Intercept[4]    -0.63      0.38    -1.22     0.61 1.02      171       30
````
````

```
TypeDere         -0.81       0.24     -1.29     -0.35 1.00
   4530     4890
Further Distributional Parameters:
     Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_
        ESS
disc     1.00      0.00     1.00     1.00    NA       NA
         NA
Draws were sampled using sampling(NUTS). For each parameter,
   Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is
   the potential
scale reduction factor on split chains (at convergence, Rhat
   = 1).
```{r}
em3 = emmeans::emmeans(m12, specs=pairwise~Type)
em3$contrasts %>%
  summary(infer=T)
```
 contrast   estimate lower.HPD upper.HPD
 De dicto - De re   0.81   0.357   1.29
Note: contrasts are still on the probit scale
Point estimate displayed: median
HPD interval probability: 0.95
```

Difference in context A (non-lawyer)

```
```{r}
all_data_NonLawyersA <- all_data_NonLawyers[which(all_data_
   NonLawyers$Context == "A"),]
all_data_NonLawyersA$Type = factor(all_data_NonLawyersA$Type,
    levels = c("De dicto", "De re"))
prior_rating <- c(brms::set_prior("normal(0,2)",class="b"),
                  brms::set_prior("normal(0,2)",class="
                     Intercept"))
                  # set_prior("lkj(2)", class = "cor"))
m100 = brms::brm(Response ~Type+(1|ParticipantID)+(1|Context)
   ,
         data=all_data_NonLawyersA,
         family = brms::cumulative(link ="probit",
         threshold="flexible"),
         init=0,
         prior=prior_rating,
         iter=4000,
         cores=2)
summary(m100)
```
Family: cumulative
  Links: mu = probit; disc = identity
```

```
Formula: Response ~ Type + (1 | ParticipantID) + (1 | Context)
   Data: all_data_NonLawyersA (Number of observations: 28)
  Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
         total post-warmup draws = 8000
Multilevel Hyperparameters:
~Context (Number of levels: 1)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     2.06      1.83     0.08     6.81 1.00     3127     3371
~ParticipantID (Number of levels: 28)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     4.42      2.20     1.32     9.60 1.00     1509     1951
Regression Coefficients:
             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept[1]    -2.38      1.31    -5.09     0.12 1.00     3411     4501
Intercept[2]    -1.98      1.25    -4.51     0.42 1.00     3681     4685
Intercept[3]    -1.30      1.20    -3.69     1.12 1.00     3876     4491
Intercept[4]     1.14      1.43    -1.33     4.32 1.00     2821     4969
TypeDere        -1.34      1.37    -4.19     1.38 1.00     2870     3658
Further Distributional Parameters:
     Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
disc     1.00      0.00     1.00     1.00   NA       NA       NA

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```{r}
em100 = emmeans::emmeans(m100, specs=pairwise~Type)
em100$contrasts %>%
  summary(infer=T)
```
 contrast   estimate lower.HPD upper.HPD
```

```
 De dicto - De re      1.32      -1.41       4.15
Note: contrasts are still on the probit scale
Point estimate displayed: median
HPD interval probability: 0.95
```

Difference in context B (non-lawyer)

```
```{r}
all_data_NonLawyersB <- all_data_NonLawyers[which(all_data_
   NonLawyers$Context == "B"),]
all_data_NonLawyersB$Type = factor(all_data_NonLawyersB$Type,
    levels = c("De dicto", "De re"))
prior_rating <- c(brms::set_prior("normal(0,2)",class="b"),
                 brms::set_prior("normal(0,2)",class="
                    Intercept"))
                 # set_prior("lkj(2)", class = "cor"))
m101 = brms::brm(Response ~Type+(1|ParticipantID)+(1|Context)
   ,
          data=all_data_NonLawyersB,
          family = brms::cumulative(link ="probit",
          threshold="flexible"),
          init=0,
          prior=prior_rating,
          iter=4000,
          cores=2)
summary(m101)
```
 Family: cumulative
  Links: mu = probit; disc = identity
Formula: Response ~ Type + (1 | ParticipantID) + (1 | Context
)
   Data: all_data_NonLawyersB (Number of observations: 28)
  Draws: 4 chains, each with iter = 4000; warmup = 2000; thin
      = 1;
         total post-warmup draws = 8000
Multilevel Hyperparameters:
~Context (Number of levels: 1)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_
                 ESS Tail_ESS
sd(Intercept)   3.45   2.46   0.36   9.69 1.01
   332   347
~ParticipantID (Number of levels: 28)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_
                 ESS Tail_ESS
sd(Intercept)   3.54   1.62   0.90   7.23 1.00
   815   1504
Regression Coefficients:
```

```
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept[1]    -1.96      1.32    -4.72     0.39 1.01      796     2057
Intercept[2]    -0.95      1.21    -3.39     1.40 1.00     1241     1322
Intercept[3]    -0.18      1.22    -2.52     2.18 1.00     1630     1981
Intercept[4]     1.71      1.50    -0.98     4.80 1.01      384      358
TypeDere         0.33      1.25    -2.22     2.78 1.00     1036     2874
Further Distributional Parameters:
     Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
disc     1.00      0.00     1.00     1.00   NA       NA       NA
Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```{r}
em101 = emmeans::emmeans(m101, specs=pairwise~Type)
em101$contrasts %>%
  summary(infer=T)
```
 contrast   estimate lower.HPD upper.HPD
 De dicto - De re   -0.324   -2.89   2.1
Note: contrasts are still on the probit scale
Point estimate displayed: median
HPD interval probability: 0.95
```

Difference in context C (non-lawyer)

```
```{r}
all_data_NonLawyersC <- all_data_NonLawyers[which(all_data_NonLawyers$Context == "C"),]
all_data_NonLawyersC$Type = factor(all_data_NonLawyersC$Type, levels = c("De dicto", "De re"))
prior_rating <- c(brms::set_prior("normal(0,2)",class="b"),
                  brms::set_prior("normal(0,2)",class="Intercept"))
                  # set_prior("lkj(2)", class = "cor"))
m102 = brms::brm(Response ~Type+(1|ParticipantID)+(1|Context),
         data=all_data_NonLawyersC,
```
```

````
          family = brms::cumulative(link ="probit",
          threshold="flexible"),
          init=0,
          prior=prior_rating,
          iter=4000,
          cores=2)
summary(m102)
```
 Family: cumulative
  Links: mu = probit; disc = identity
Formula: Response ~ Type + (1 | ParticipantID) + (1 | Context)
   Data: all_data_NonLawyersC (Number of observations: 28)
  Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
         total post-warmup draws = 8000
Multilevel Hyperparameters:
~Context (Number of levels: 1)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     2.88      2.38     0.17     9.35 1.00     1029      383
~ParticipantID (Number of levels: 28)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     2.75      1.66     0.27     6.87 1.00      829      397
Regression Coefficients:
             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept[1]    -3.04      1.37    -5.85    -0.49 1.00     3102     4462
Intercept[2]    -1.86      1.19    -4.20     0.50 1.00     2102      757
Intercept[3]    -1.54      1.17    -3.79     0.74 1.00     2792     1772
Intercept[4]    -0.14      1.27    -2.41     2.57 1.00     3126     4000
TypeDere        -2.32      1.20    -4.81    -0.10 1.00     1611      674
Further Distributional Parameters:
     Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
disc     1.00      0.00     1.00     1.00   NA       NA       NA
Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
````
````

```
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```{r}
em102 = emmeans::emmeans(m102, specs=pairwise~Type)
em102$contrasts %>%
  summary(infer=T)
```
 contrast   estimate lower.HPD upper.HPD
 De dicto - De re   2.25   -0.0481   4.64

Note: contrasts are still on the probit scale
Point estimate displayed: median
HPD interval probability: 0.95
```

Difference in context D (non-lawyer)

```
```{r}
all_data_NonLawyersD <- all_data_NonLawyers[which(all_data_NonLawyers$Context == "D"),]
all_data_NonLawyersD$Type = factor(all_data_NonLawyersD$Type, levels = c("De dicto", "De re"))
prior_rating <- c(brms::set_prior("normal(0,2)",class="b"),
                brms::set_prior("normal(0,2)",class="Intercept"))
                # set_prior("lkj(2)", class = "cor"))
m103 = brms::brm(Response ~Type+(1|ParticipantID)+(1|Context),
         data=all_data_NonLawyersD,
         family = brms::cumulative(link ="probit",
         threshold="flexible"),
         init=0,
         prior=prior_rating,
         iter=4000,
         cores=2)
summary(m103)
```
 Family: cumulative
  Links: mu = probit; disc = identity
Formula: Response ~ Type + (1 | ParticipantID) + (1 | Context)
   Data: all_data_NonLawyersD (Number of observations: 28)
  Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
         total post-warmup draws = 8000
Multilevel Hyperparameters:
~Context (Number of levels: 1)
```

````
                Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     2.74      2.41     0.16     9.26 1.00     1708      942
~ParticipantID (Number of levels: 28)
                Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     2.73      1.57     0.36     6.46 1.01      828     1173
Regression Coefficients:
               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept[1]     -3.24      1.40    -6.13    -0.58 1.00     1832     1769
Intercept[2]     -2.51      1.27    -5.01    -0.00 1.00     2492     2161
Intercept[3]     -1.91      1.24    -4.39     0.54 1.00     2727     2158
Intercept[4]     -0.57      1.31    -2.95     2.17 1.00     2986     3039
TypeDere         -2.59      1.24    -5.17    -0.26 1.00     2865     3365
Further Distributional Parameters:
     Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
disc     1.00      0.00     1.00     1.00   NA       NA       NA
Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```{r}
em103 = emmeans::emmeans(m103, specs=pairwise~Type)
em103$contrasts %>%
  summary(infer=T)
```
 contrast          estimate lower.HPD upper.HPD
 De dicto - De re      2.54     0.205       5.1

Note: contrasts are still on the probit scale
Point estimate displayed: median
HPD interval probability: 0.95
````

Difference between type among lawyers

````
```{r}
````
````

````
all_data_Lawyers <- all_data[which(all_data$Group == "Lawyer"),]
all_data_Lawyers$Type = factor(all_data_Lawyers$Type, levels = c("De dicto", "De re"))
prior_rating <- c(brms::set_prior("normal(0,2)",class="b"),
                  brms::set_prior("normal(0,2)",class="Intercept"))
                  # set_prior("lkj(2)", class = "cor"))
m13 = brms::brm(Response ~Type+(1|ParticipantID)+(1|Context),
         data=all_data_Lawyers,
         family = brms::cumulative(link ="probit",
         threshold="flexible"),
         init=0,
         prior=prior_rating,
         iter=4000,
         cores=2)
summary(m13)
```
 Family: cumulative
  Links: mu = probit; disc = identity
Formula: Response ~ Type + (1 | ParticipantID) + (1 | Context)
   Data: all_data_Lawyers (Number of observations: 72)
  Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
         total post-warmup draws = 8000
Multilevel Hyperparameters:
~Context (Number of levels: 4)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     0.93      0.59     0.25     2.49 1.00     2407     2746
~ParticipantID (Number of levels: 18)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     0.48      0.29     0.03     1.11 1.00     2083     3290
Regression Coefficients:
             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept[1]    -1.45      0.52    -2.42    -0.36 1.00     3153     3218
Intercept[2]    -1.12      0.51    -2.06    -0.05 1.00     3131     3243
Intercept[3]    -1.02      0.51    -1.95     0.04 1.00     3125     3192
Intercept[4]    -0.39      0.50    -1.33     0.69 1.00     3247     3014
````
````

```
TypeDere        -0.87      0.31    -1.48    -0.27 1.00     5644     4140
Further Distributional Parameters:
     Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
disc     1.00      0.00     1.00     1.00   NA       NA       NA
Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```{r}
em4 = emmeans::emmeans(m13, specs=pairwise~Type)
em4$contrasts %>%
  summary(infer=T)
```
 contrast   estimate lower.HPD upper.HPD
 De dicto - De re   0.862   0.261   1.47
Note: contrasts are still on the probit scale
Point estimate displayed: median
HPD interval probability: 0.95
```

Difference in context A (lawyer)

```
```{r}
all_data_LawyersA <- all_data_Lawyers[which(all_data_Lawyers$Context == "A"),]
all_data_LawyersA$Type = factor(all_data_LawyersA$Type, levels = c("De dicto", "De re"))
prior_rating <- c(brms::set_prior("normal(0,2)",class="b"),
                  brms::set_prior("normal(0,2)",class="Intercept"))
                  # set_prior("lkj(2)", class = "cor"))
m200 = brms::brm(Response ~Type+(1|ParticipantID)+(1|Context),
         data=all_data_LawyersA,
         family = brms::cumulative(link ="probit",
         threshold="flexible"),
         init=0,
         prior=prior_rating,
         iter=4000,
         cores=2)
summary(m200)
```
 Family: cumulative
  Links: mu = probit; disc = identity
```

```
Formula: Response ~ Type + (1 | ParticipantID) + (1 | Context)
   Data: all_data_LawyersA (Number of observations: 18)
  Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
         total post-warmup draws = 8000
Multilevel Hyperparameters:
~Context (Number of levels: 1)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     1.93      2.02     0.06     7.13 1.00 1105      343
~ParticipantID (Number of levels: 18)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     5.44      2.63     1.76    11.84 1.00 1789     2086
Regression Coefficients:
             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept[1]    -1.96      1.40    -4.93     0.55 1.00 3233     4024
Intercept[2]    -0.49      1.24    -2.95     1.87 1.00 3970     5170
Intercept[3]    -0.02      1.24    -2.46     2.38 1.00 4247     5467
Intercept[4]     1.98      1.45    -0.70     5.03 1.00 4515     5255
TypeDere        -0.61      1.59    -3.75     2.44 1.00 2788     5105
Further Distributional Parameters:
     Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
disc     1.00      0.00     1.00     1.00   NA       NA       NA
Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```{r}
em200 = emmeans::emmeans(m200, specs=pairwise~Type)
em200$contrasts %>%
  summary(infer=T)
```
 contrast   estimate lower.HPD upper.HPD
 De dicto - De re   0.601   -2.51   3.68
```

````
Note: contrasts are still on the probit scale
Point estimate displayed: median
HPD interval probability: 0.95
````

Difference in context B (lawyer)

````
```{r}
all_data_LawyersB <- all_data_Lawyers[which(all_data_Lawyers$Context == "B"),]
all_data_LawyersB$Type = factor(all_data_LawyersB$Type, levels = c("De dicto", "De re"))
prior_rating <- c(brms::set_prior("normal(0,2)",class="b"),
                  brms::set_prior("normal(0,2)",class="Intercept"))
                  # set_prior("lkj(2)", class = "cor"))
m201 = brms::brm(Response ~Type+(1|ParticipantID)+(1|Context),
         data=all_data_LawyersB,
         family = brms::cumulative(link ="probit",
         threshold="flexible"),
         init=0,
         prior=prior_rating,
         iter=4000,
         cores=2)
summary(m201)
```
 Family: cumulative
  Links: mu = probit; disc = identity
Formula: Response ~ Type + (1 | ParticipantID) + (1 | Context)
   Data: all_data_LawyersB (Number of observations: 18)
  Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
         total post-warmup draws = 8000
Multilevel Hyperparameters:
~Context (Number of levels: 1)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     3.50      2.74     0.29    10.76 1.01 352      156
~ParticipantID (Number of levels: 18)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     2.75      1.87     0.25     8.12 1.02 303       82
Regression Coefficients:
             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
````

```
Intercept[1]    -1.21      1.31    -3.85     1.32 1.00     2060     3649
Intercept[2]    -0.53      1.25    -2.95     1.98 1.00     2021     4022
Intercept[3]     0.50      1.29    -1.96     3.11 1.00     3206     3823
Intercept[4]     2.06      1.48    -0.56     5.21 1.00     2416     3043
TypeDere         1.50      1.28    -0.99     4.08 1.00     1656     2529
Further Distributional Parameters:
     Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
disc     1.00      0.00     1.00     1.00   NA       NA       NA
Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```{r}
em201 = emmeans::emmeans(m201, specs=pairwise~Type)
em201$contrasts %>%
  summary(infer=T)
```
 contrast   estimate lower.HPD upper.HPD
 De dicto - De re   -1.47   -4.06   1
Note: contrasts are still on the probit scale
Point estimate displayed: median
HPD interval probability: 0.95
```

Difference in context C (lawyer)

```
```{r}
all_data_LawyersC <- all_data_Lawyers[which(all_data_Lawyers$Context == "C"),]
all_data_LawyersC$Type = factor(all_data_LawyersC$Type, levels = c("De dicto", "De re"))
prior_rating <- c(brms::set_prior("normal(0,2)",class="b"),
                brms::set_prior("normal(0,2)",class="Intercept"))
                # set_prior("lkj(2)", class = "cor"))
m202 = brms::brm(Response ~Type+(1|ParticipantID)+(1|Context),
         data=all_data_LawyersC,
         family = brms::cumulative(link ="probit",
         threshold="flexible"),
```
```

````
         init=0,
         prior=prior_rating,
         iter=4000,
         cores=2)
summary(m202)
```
 Family: cumulative
  Links: mu = probit; disc = identity
Formula: Response ~ Type + (1 | ParticipantID) + (1 | Context)
   Data: all_data_LawyersC (Number of observations: 18)
  Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
         total post-warmup draws = 8000
Multilevel Hyperparameters:
~Context (Number of levels: 1)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     2.07      1.69     0.10     6.64 1.00     2641     1451
~ParticipantID (Number of levels: 18)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     1.28      1.03     0.06     3.90 1.00     1399     2497
Regression Coefficients:
             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept[1]    -3.55      1.29    -6.20    -1.11 1.00     3853     4332
Intercept[2]    -2.59      1.15    -4.84    -0.35 1.00     5330     5818
Intercept[3]    -2.30      1.13    -4.54    -0.03 1.00     5695     5463
Intercept[4]    -1.01      1.14    -3.18     1.32 1.00     4946     3580
TypeDere        -3.61      1.17    -6.14    -1.47 1.00     3942     2231
Further Distributional Parameters:
     Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
disc     1.00      0.00     1.00     1.00   NA       NA       NA
Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
````
````

```
scale reduction factor on split chains (at convergence, Rhat = 1).
```{r}
em202 = emmeans::emmeans(m202, specs=pairwise~Type)
em202$contrasts %>%
  summary(infer=T)
```
 contrast   estimate lower.HPD upper.HPD
 De dicto - De re   3.53   1.31   5.94
Note: contrasts are still on the probit scale
Point estimate displayed: median
HPD interval probability: 0.95
```

Difference in context D (lawyer)

```
```{r}
all_data_LawyersD <- all_data_Lawyers[which(all_data_Lawyers$Context == "D"),]
all_data_LawyersD$Type = factor(all_data_LawyersD$Type, levels = c("De dicto", "De re"))
prior_rating <- c(brms::set_prior("normal(0,2)",class="b"),
                  brms::set_prior("normal(0,2)",class="Intercept"))
                  # set_prior("lkj(2)", class = "cor"))
m203 = brms::brm(Response ~Type+(1|ParticipantID)+(1|Context),
         data=all_data_LawyersD,
         family = brms::cumulative(link ="probit",
         threshold="flexible"),
         init=0,
         prior=prior_rating,
         iter=4000,
         cores=2)
summary(m203)
```
 Family: cumulative
  Links: mu = probit; disc = identity
Formula: Response ~ Type + (1 | ParticipantID) + (1 | Context)
   Data: all_data_LawyersD (Number of observations: 18)
  Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
         total post-warmup draws = 8000
Multilevel Hyperparameters:
~Context (Number of levels: 1)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
```

```
sd(Intercept)     2.73      2.44     0.14     8.97 1.00    1548     615
~ParticipantID (Number of levels: 18)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     2.44      1.56     0.22     6.30 1.00    1194    2008
Regression Coefficients:
             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept[1]    -3.05      1.36    -5.88    -0.51 1.00    2978    2746
Intercept[2]    -2.09      1.23    -4.58     0.28 1.00    4282    5320
Intercept[3]    -1.67      1.22    -4.05     0.69 1.00    4168    4906
Intercept[4]     0.06      1.34    -2.34     2.88 1.00    3542    4403
TypeDere        -2.35      1.27    -4.96     0.11 1.00    3016    3600
Further Distributional Parameters:
     Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
disc     1.00      0.00     1.00     1.00   NA       NA       NA
Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```{r}
em203 = emmeans::emmeans(m203, specs=pairwise~Type)
em203$contrasts %>%
  summary(infer=T)
```
 contrast   estimate lower.HPD upper.HPD
 De dicto - De re   2.31   -0.106   4.97
Note: contrasts are still on the probit scale
Point estimate displayed: median
HPD interval probability: 0.95
```

Detailed Code and Results for Ordinal Regression Model of Judgment Task

Difference by type among non-lawyers

```
```{r}
```
```

```
all_data_NonLawyers$Response = as.factor(all_data_NonLawyers$Response)
m20 = ordinal::clmm(Response ~Type+(1|ParticipantID)+(1|Context),
         data=all_data_NonLawyers)
summary(m20)
```
Cumulative Link Mixed Model fitted with the Laplace approximation
formula: Response ~ Type + (1 | ParticipantID) + (1 | Context)
data: all_data_NonLawyers
Random effects:
 Groups   Name   Variance Std.Dev.
 ParticipantID (Intercept) 1.50e-14 1.225e-07
 Context   (Intercept) 1.53e-02 1.237e-01
Number of groups: ParticipantID 28, Context 4
Coefficients:
          Estimate Std. Error z value Pr(>|z|)
TypeDe re -1.2672   0.4003 -3.166 0.00155 **

Signif. codes: 0   ***   0.001   **   0.01   *   0.05   .   0.1   1
Threshold coefficients:
    Estimate Std. Error z value
1|2 -2.5579   0.3907 -6.547
2|3 -2.2748   0.3703 -6.143
3|4 -2.0891   0.3585 -5.827
4|5 -1.1379   0.3133 -3.632
```

**Difference by type among lawyers**

```
```{r}
all_data_Lawyers$Response = as.factor(all_data_Lawyers$Response)
m15 = ordinal::clmm(Response ~Type+(1|ParticipantID)+(1|Context),
         data=all_data_Lawyers)
summary(m15)
```
Cumulative Link Mixed Model fitted with the Laplace approximation
formula: Response ~ Type + (1 | ParticipantID) + (1 | Context)
data: all_data_Lawyers
Random effects:
 Groups   Name   Variance Std.Dev.
 ParticipantID (Intercept) 0.2522   0.5022
```
```

```
 Context       (Intercept) 0.5894   0.7677
Number of groups:  ParticipantID 18,  Context 4
Coefficients:
          Estimate Std. Error z value Pr(>|z|)
TypeDe re  -1.3552     0.5193  -2.609  0.00907 **
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1
Threshold coefficients:
    Estimate Std. Error z value
1|2  -2.4448     0.6680  -3.660
2|3  -1.9639     0.6342  -3.097
3|4  -1.8785     0.6287  -2.988
4|5  -0.8933     0.5712  -1.564
```

Difference by between lawyers and non-lawyers in *de re*

```
```{r}
all_data_Re <- all_data[which(all_data$Type == "De re"),]
all_data_Re$Response = as.factor(all_data_Re$Response)
m21 = ordinal::clmm(Response ~Group+(1|ParticipantID)+(1|Context),
         data=all_data_Re)
summary(m21)
```
Cumulative Link Mixed Model fitted with the Laplace approximation
formula: Response ~ Group + (1 | ParticipantID) + (1 | Context)
data: all_data_Re
Random effects:
 Groups   Name   Variance Std.Dev.
 ParticipantID (Intercept) 0.4402   0.6634
 Context   (Intercept) 0.6962   0.8344
Number of groups: ParticipantID 46, Context 4
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
GroupLawyer -0.3457   0.4646 -0.744   0.457
Threshold coefficients:
    Estimate Std. Error z value
1|2 -1.4989   0.5735 -2.614
2|3 -1.1124   0.5505 -2.021
3|4 -0.9921   0.5448 -1.821
4|5   0.1130   0.5230   0.216
```

Difference by between lawyers and non-lawyers in *de dicto*

```
```{r}
```
```

````
all_data_Dicto <- all_data[which(all_data$Type == "De dicto"),]
all_data_Dicto$Response = as.factor(all_data_Dicto$Response)
m22 = ordinal::clmm(Response ~Group+(1|ParticipantID)+(1|Context),
         data=all_data_Dicto)
summary(m22)
```
Cumulative Link Mixed Model fitted with the Laplace approximation
formula: Response ~ Group + (1 | ParticipantID) + (1 | Context)
data:    all_data_Dicto
Random effects:
 Groups        Name        Variance Std.Dev.
 ParticipantID (Intercept) 1.288    1.135
 Context       (Intercept) 1.399    1.183
Number of groups:  ParticipantID 46,  Context 4
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
GroupLawyer  -0.5945     0.6641  -0.895    0.371
Threshold coefficients:
    Estimate Std. Error z value
1|2  -3.8106     1.0815  -3.524
2|3  -3.3001     1.0089  -3.271
3|4  -3.0227     0.9726  -3.108
4|5  -1.7347     0.8174  -2.122
````

Detailed Code and Results for Linear Mixed Effects Model for Response Duration

Is a response significantly quicker or slower?

````
```{r}
m1000 = lmerTest::lmer(Response.Duration ~Response + (1|ParticipantID) + (1|Context), data=cleaned_all_dataRD)
summary(m1000)
```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [lmerModLmerTest]
Formula: Response.Duration ~ Response + (1 | ParticipantID) + (1 | Context)
   Data: cleaned_all_dataRD
REML criterion at convergence: 3292.7
Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.1267 -0.4091 -0.3078 -0.1028  4.6128
Random effects:
 Groups        Name        Variance Std.Dev.
````
````

```
 ParticipantID (Intercept) 1647253  1283
 Context       (Intercept)       0     0
 Residual                  9756224  3123
Number of obs: 177, groups:  ParticipantID, 46; Context, 4
Fixed effects:
                          Estimate Std. Error     df t value Pr(>|t|)
(Intercept)                 2380.7      364.4   66.9   6.534 1.03e-08 ***
ResponseHighly disagree     -389.3      670.5  153.5  -0.581    0.562
ResponseSomewhat agree       119.6      698.1  169.3   0.171    0.864
ResponseSomewhat disagree   -627.9     1301.5  170.3  -0.482    0.630
ResponseUncertain           2344.4     1923.8  163.4   1.219    0.225
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1    1
Correlation of Fixed Effects:
            (Intr) RspnHd RspnSa RspnSd
RspnsHghlyd -0.402
RspnsSmwhta -0.427  0.209
RspnsSmwhtd -0.225  0.107  0.133
RspnsUncrtn -0.142  0.069  0.063  0.080
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

Response duration by type overall

```
```{r}
m500 = lmerTest::lmer(Response.Duration ~ Type + (1|ParticipantID) + (1|Context), data=cleaned_all_dataRD)
summary(m500)
```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [lmerModLmerTest]
Formula: Response.Duration ~ Type + (1 | ParticipantID) + (1 | Context)
   Data: cleaned_all_dataRD
REML criterion at convergence: 3339.6
Scaled residuals:
    Min   1Q Median   3Q   Max
-1.1579 -0.4846 -0.2331 -0.0533 4.4886
Random effects:
 Groups   Name   Variance Std.Dev.
 ParticipantID (Intercept) 1774983 1332
```

```
 Context       (Intercept)       0     0
 Residual                  9379746  3063
Number of obs: 177, groups:  ParticipantID, 46; Context, 4
Fixed effects:
             Estimate Std. Error      df t value Pr(>|t|)
(Intercept)   2832.60     383.51   92.42   7.386 6.54e-11 ***
TypeDe dicto  -940.25     461.24  124.64  -2.039   0.0436 *
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05
      .    0.1         1
Correlation of Fixed Effects:
            (Intr)
TypeDedicto -0.613
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

Response duration by type in context A

```
```{r}
cleaned_all_dataRDA <- cleaned_all_dataRD[cleaned_all_dataRD$
   Context == "Story A", ]
cleaned_all_dataRDA$Response=factor(cleaned_all_dataRDA$
   Response, levels = c("Uncertain", "Highly disagree", "
   Somewhat disagree", "Somewhat agree", "Highly agree"))
m501 = lmerTest::lmer(Response.Duration ~ Type + (1|Group),
   data=cleaned_all_dataRDA)
summary(m501)
```
Linear mixed model fit by REML. t-tests use Satterthwaite's
   method [lmerModLmerTest]
Formula: Response.Duration ~ Type + (1 | Group)
   Data: cleaned_all_dataRDA
REML criterion at convergence: 800.3
Scaled residuals:
    Min   1Q Median   3Q   Max
-1.0993 -0.4176 -0.1856 0.0925 4.1167
Random effects:
 Groups   Name   Variance Std.Dev.
 Group   (Intercept) 393613   627.4
 Residual   6003620 2450.2
Number of obs: 45, groups: Group, 2
Fixed effects:
              Estimate Std. Error   df t value Pr(>|t|)
(Intercept)   2897.926   687.308   1.949   4.216   0.0543
   .
TypeDe dicto -1787.363   730.768   42.011 -2.446   0.0187
   *

```

```
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1
Correlation of Fixed Effects:
            (Intr)
TypeDedicto -0.542
```

Response duration by type in context B

```
‘‘‘{r}
cleaned_all_dataRDB <- cleaned_all_dataRD[cleaned_all_dataRD$Context == "Story B", ]
cleaned_all_dataRDB$Response=factor(cleaned_all_dataRDB$Response, levels = c("Uncertain", "Highly disagree", "Somewhat disagree", "Somewhat agree", "Highly agree"))
m502 = lmerTest::lmer(Response.Duration ~ Type + (1|Group), data=cleaned_all_dataRDB)
summary(m502)
‘‘‘
Linear mixed model fit by REML. t-tests use Satterthwaite’s method [lmerModLmerTest]
Formula: Response.Duration ~ Type + (1 | Group)
   Data: cleaned_all_dataRDB
REML criterion at convergence: 802.4
Scaled residuals:
    Min      1Q  Median      3Q     Max
-0.6966 -0.4967 -0.2739 -0.0986  3.9111
Random effects:
 Groups   Name        Variance Std.Dev.
 Group    (Intercept)        0    0
 Residual             10005806 3163
Number of obs: 44, groups:  Group, 2
Fixed effects:
             Estimate Std. Error     df t value Pr(>|t|)
(Intercept)    1470.3      690.3   42.0    2.13   0.0391 *
TypeDe dicto    907.1      954.7   42.0    0.95   0.3475
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1
Correlation of Fixed Effects:
            (Intr)
TypeDedicto -0.723
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help(’isSingular’)
```

Response duration by type in context C

```
‘‘‘{r}
cleaned_all_dataRDC <- cleaned_all_dataRD[cleaned_all_dataRD$Context == "Story C", ]
```

````
cleaned_all_dataRDC$Response=factor(cleaned_all_dataRDC$Response, levels = c("Uncertain", "Highly disagree", "Somewhat disagree", "Somewhat agree", "Highly agree"))
m503 = lmerTest::lmer(Response.Duration ~ Type + (1|Group), data=cleaned_all_dataRDC)
summary(m503)
```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [lmerModLmerTest]
Formula: Response.Duration ~ Type + (1 | Group)
   Data: cleaned_all_dataRDC
REML criterion at convergence: 807.4
Scaled residuals:
    Min      1Q  Median      3Q     Max
-0.8153 -0.5414 -0.3614 -0.0750  3.9321
Random effects:
 Groups   Name        Variance Std.Dev.
 Group    (Intercept)        0    0
 Residual             11289314 3360
Number of obs: 44, groups:  Group, 2
Fixed effects:
             Estimate Std. Error     df t value Pr(>|t|)
(Intercept)    3050.4      733.2   42.0   4.160 0.000154 ***
TypeDe dicto  -1071.5     1014.1   42.0  -1.057 0.296745
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1
Correlation of Fixed Effects:
            (Intr)
TypeDedicto -0.723
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
````

Response duration by type in context D

````
```{r}
cleaned_all_dataRDD <- cleaned_all_dataRD[cleaned_all_dataRD$Context == "Story D", ]
cleaned_all_dataRDD$Response=factor(cleaned_all_dataRDD$Response, levels = c("Uncertain", "Highly disagree", "Somewhat disagree", "Somewhat agree", "Highly agree"))
m504 = lmerTest::lmer(Response.Duration ~ Type + (1|Group), data=cleaned_all_dataRDD)
summary(m504)
```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [lmerModLmerTest]
Formula: Response.Duration ~ Type + (1 | Group)
````
````

```
   Data: cleaned_all_dataRDD
REML criterion at convergence: 824
Scaled residuals:
    Min      1Q  Median      3Q     Max
-0.8234 -0.6829 -0.3129  0.1035  3.6835
Random effects:
 Groups   Name        Variance Std.Dev.
 Group    (Intercept)        0    0
 Residual             16766365 4095
Number of obs: 44, groups:  Group, 2
Fixed effects:
             Estimate Std. Error      df t value Pr(>|t|)
(Intercept)    3601.7      853.8    42.0   4.218 0.000128 ***
TypeDe dicto  -1703.0     1235.9    42.0  -1.378 0.175518
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
Correlation of Fixed Effects:
            (Intr)
TypeDedicto -0.691
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

Are lawyers faster than non-lawyers overall?

```
```{r}
m600 = lmer(Response.Duration ~ Group + (1|ParticipantID) +
   (1|Context), data=cleaned_all_dataRD)
summary(m600)
```
Linear mixed model fit by REML. t-tests use Satterthwaite's
   method [lmerModLmerTest]
Formula: Response.Duration ~ Group + (1 | ParticipantID) + (1
    | Context)
   Data: cleaned_all_dataRD
REML criterion at convergence: 3341.8
Scaled residuals:
    Min   1Q Median   3Q   Max
-1.1903 -0.4140 -0.3131 -0.1242 4.5781
Random effects:
 Groups   Name   Variance Std.Dev.
 ParticipantID (Intercept) 1697331 1303
 Context   (Intercept)   0   0
 Residual   9594575 3098
Number of obs: 177, groups: ParticipantID, 46; Context, 4
Fixed effects:
            Estimate Std. Error   df t value Pr(>|t|)
(Intercept) 2638.32   388.48   38.15   6.791 4.63e-08 ***
```

```
GroupLawyer  -726.01     618.61   37.74  -1.174    0.248
---
Signif. codes:  0   ***   0.001   **   0.01   *   0.05
     .   0.1       1
Correlation of Fixed Effects:
            (Intr)
GroupLawyer -0.628
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

Are lawyers faster in one of the conditions than in the other?

```
```{r}
m601 = lmerTest::lmer(Response.Duration ~ Type + (1|
   ParticipantID) + (1|Context), data=cleaned_all_
   dataRDLawyer)
summary(m601)
```
Linear mixed model fit by REML. t-tests use Satterthwaite's
   method [lmerModLmerTest]
Formula: Response.Duration ~ Type + (1 | ParticipantID) + (1
   | Context)
   Data: cleaned_all_dataRDLawyer
REML criterion at convergence: 1288.4
Scaled residuals:
    Min   1Q Median   3Q   Max
-0.8925 -0.5329 -0.2269 -0.0043 4.7800
Random effects:
 Groups   Name   Variance Std.Dev.
 ParticipantID (Intercept) 8.329e+05 9.126e+02
 Context   (Intercept) 2.213e-10 1.488e-05
 Residual   8.221e+06 2.867e+03
Number of obs: 70, groups: ParticipantID, 18; Context, 4
Fixed effects:
             Estimate Std. Error   df t value Pr(>|t|)
(Intercept)   2701.00   537.62   44.05   5.024 8.88e-06 **
   *
TypeDe dicto -1528.99   686.37   52.14 -2.228   0.0302 *

Signif. codes: 0   ***   0.001   **   0.01   *   0.05
     .   0.1   1
Correlation of Fixed Effects:
            (Intr)
TypeDedicto -0.658
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

Are non-lawyers faster in one of the conditions than in the other?

```
```{r}
m602 = lmerTest::lmer(Response.Duration ~ Type + (1|
   ParticipantID) + (1|Context), data=cleaned_all_
   dataRDNonLawyer)
summary(m602)
```
Linear mixed model fit by REML. t-tests use Satterthwaite's
   method [lmerModLmerTest]
Formula: Response.Duration ~ Type + (1 | ParticipantID) + (1
   | Context)
   Data: cleaned_all_dataRDNonLawyer
REML criterion at convergence: 2016.5
Scaled residuals:
    Min   1Q Median   3Q   Max
-1.2437 -0.4463 -0.2704 -0.0384 4.2367
Random effects:
 Groups   Name   Variance Std.Dev.
 ParticipantID (Intercept) 2548493 1596
 Context   (Intercept)   0   0
 Residual   9983870 3160
Number of obs: 107, groups: ParticipantID, 28; Context, 4
Fixed effects:
             Estimate Std. Error   df t value Pr(>|t|)
(Intercept)   2932.74   529.93   49.23   5.534 1.2e-06 ***
TypeDe dicto -551.39   612.18   73.30 -0.901   0.371

Signif. codes: 0   ***   0.001   **   0.01   *   0.05
      .   0.1   1
Correlation of Fixed Effects:
            (Intr)
TypeDedicto -0.584
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

Are lawyers faster than non-lawyers in *de re*

```
```{r}
cleaned_all_dataRDre <- cleaned_all_dataRD[cleaned_all_dataRD
   $Type =="De re",]
m610 = lmerTest::lmer(Response.Duration ~ Group + (1|
   ParticipantID) + (1|Context), data=cleaned_all_dataRDre)
summary(m610)
```
Linear mixed model fit by REML. t-tests use Satterthwaite's
   method [lmerModLmerTest]
Formula: Response.Duration ~ Group + (1 | ParticipantID) + (1
    | Context)
   Data: cleaned_all_dataRDre
```

```
REML criterion at convergence: 1656.9
Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.2840 -0.3959 -0.3339 -0.0035  3.4307
Random effects:
 Groups        Name        Variance Std.Dev.
 ParticipantID (Intercept)  4547492 2132
 Context       (Intercept)        0    0
 Residual                  11831354 3440
Number of obs: 87, groups:  ParticipantID, 46; Context, 4
Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  2974.44     623.48   32.12   4.771 3.83e-05 ***
GroupLawyer  -263.82     997.15   32.17  -0.265    0.793
---
Signif. codes:  0   ***   0.001   **   0.01   *   0.05
     .   0.1        1
Correlation of Fixed Effects:
            (Intr)
GroupLawyer -0.625
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

Are lawyers faster than non-lawyers in *de dicto*

```
```{r}
cleaned_all_dataRDdicto <- cleaned_all_dataRD[cleaned_all_
   dataRD$Type =="De dicto",]
m611 = lmerTest::lmer(Response.Duration ~ Group + (1|
   ParticipantID) + (1|Context), data=cleaned_all_dataRDdicto
   )
summary(m611)
```
Linear mixed model fit by REML. t-tests use Satterthwaite's
   method [lmerModLmerTest]
Formula: Response.Duration ~ Group + (1 | ParticipantID) + (1
    | Context)
   Data: cleaned_all_dataRDdicto

REML criterion at convergence: 1633.8
Scaled residuals:
    Min   1Q Median   3Q   Max
-0.7559 -0.5704 -0.2956 -0.0735 4.9879
Random effects:
 Groups   Name   Variance Std.Dev.
 ParticipantID (Intercept)   0   0
 Context   (Intercept)   0   0
 Residual   6212175 2492
```

```
Number of obs: 90, groups:  ParticipantID, 46; Context, 4
Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)   2317.0      339.2    88.0   6.831 1.05e-09 ***
GroupLawyer  -1145.0      536.3    88.0  -2.135   0.0355 *
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05
      .    0.1          1
Correlation of Fixed Effects:
            (Intr)
GroupLawyer -0.632
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```