

Norm Conflicts and Epistemic Modals

Niels Skovgaard-Olsen

University of Freiburg

John Cantwell

KTH Royal Institute of Technology

Author Note

Niels Skovgaard-Olsen, Department of Social Psychology and Methodology, University of Freiburg, Germany. John Cantwell, KTH Royal Institute of Technology, Stockholm, Sweden.

Correspondence concerning this article should be addressed to Niels Skovgaard-Olsen (niels.skovgaard.olsen@psychologie.uni-freiburg.de, n.s.olsen@gmail.com).

The supplemental materials including stimulus materials, screen shots, data, and R-scripts have been made available on the *osf* project page: <https://osf.io/g6vym/>.

Acknowledgement: This research was supported by a research grant (497332489) to Niels Skovgaard-Olsen from the German Research Council (DFG).

Abstract

Statements containing epistemic modals (e.g., “by spring 2023 most European countries may have the Covid-19 pandemic under control”) are common expressions of epistemic uncertainty. In this paper, previous published findings (Knobe & Yalcin, 2014; Khoo & Phillips, 2018) on the opposition between Contextualism and Relativism for epistemic modals are re-examined. It is found that these findings contain a substantial degree of individual variation. To investigate whether participants differ in their interpretation of epistemic modals, an experiment with multiple phases and sessions is used to classify participants according to the three semantic theories of Relativism, Contextualism, and Objectivism. Through this study, some of the first empirical evidence for the kind of truth-value shifts postulated by semantic Relativism is presented. It is furthermore found that participants’ disagreement judgments match their truth evaluations and that participants are capable of distinguishing between truth and justification. In a second experimental session, it is investigated whether participants thus classified follow the norm of retraction which Relativism uses to account for argumentation with epistemic modals. Here the results are less favorable for Relativism. In a second experiment, these results are replicated and the normative beliefs of participants concerning the norm of retraction are investigated following work on measuring norms by Bicchieri (2017). Again, it is found that on average participants show no strong preferences concerning the norm of retraction for epistemic modals. Yet, it was found that participants who had committed to Objectivism and had training in logics applied the norm of retraction to might-statements. These results present a substantial challenge to the account of argumentation with epistemic modals presented in MacFarlane (2014), as discussed.

Keywords: Epistemic Modals, Relativism, Norm Conflicts, Retraction, Semantics, Truth Conditions, Argumentation.

Norm Conflicts and Epistemic Modals

Epistemic modals are a collection of linguistic expressions primarily used to express varying degrees of certainty, uncertainty or ignorance, which concern possibilities that are not excluded by what is known (Portner, 2009; Lassiter, 2017). Typical examples are might-modals, as in “in 2023 the status of Covid-19 might shift from pandemic to endemic”, must-modals, as in “from the rapidity of its spread, omicron must be more infectious than delta”, and various modalities for expressing likelihood or probability, as in “it is likely that omicron leads to less severe hospitalizations”. While their use is ubiquitous in all forms of discourse, they have proven something of a conundrum for theoreticians trying to account for their meaning, raising a host of fundamental issues relating to objectivity, subjectivity, truth-relativity, and context-dependence in discourse.

Traditionally, these issues have been investigated in linguistics and philosophy. Recently, psychologists have taken an interest in epistemic modality as well, with some arguing that epistemic possibilities should be made the foundation of psychology of reasoning through mental model theory (Hinterecker et al. 2016; Johnson-Laird & Ragni, 2019), and other psychologists taking a more critical perspective (Oaksford et al., 2019; Over, 2022).

Occasionally, these semantic issues make their appearance in the political discourse. A good example is when Dr. Fauci in March 2020 made the following statement about the use of the anti-malarial drug hydroxychloroquine for the treatment of Covid-19: “What I’m saying is that it might — it might be effective. I’m not saying that it isn’t. It might be effective.”¹ As of July 2020, Dr. Fauci concluded that all of the randomized, controlled clinical trials had consistently shown that hydroxychloroquine was not effective against Covid-19. The different semantic views reviewed below (Contextualism, Relativism, and

¹ www.washingtonpost.com

Objectivism), differ in their truth evaluations and on whether this correction implies that the previously asserted might-statement should later be retracted, as illustrated in Table 1.

Table 1. Norm Conflict with Epistemic Modals

	Contextualism	Relativism	Objectivism
Was Dr. Fauci's statement true at 03.2020?	Yes	Yes	No
Is Dr. Fauci's statement true in 07.2020?	Yes	No	No
Should the might statement be retracted?	No	Yes	Yes
Is Dr. Fauci at fault for making the assertion?	No	No	No

Note. The table displays evaluations of a might statement as uttered at 03.2020.

None of the views hold that Dr. Fauci was at fault for making the assertions that he did. Indeed, he was warranted in making the claim in March 2020 and did everything that he was supposed to. The question at issue is just what implications our now improved information state has for the truth evaluations of his claims. Looking back, we can then separate truth evaluations of his claim given the available evidence in March 2020 (context c_1) and in July 2020 (c_2). For questions of epistemic warrant only c_1 is relevant, but when trying to determine for ourselves whether what Dr. Fauci said was true, Relativism and Objectivism would use the improved information state in c_2 to reach a negative verdict. Below we will elaborate on how the different judgments arise. For the moment, we just present them as an exhibit of how differences in semantic interpretation can affect the normative reactions to everyday events.

Through our experiments, we investigate whether the norm conflict highlighted by Table 1 maps onto individual variation in the semantic interpretations of epistemic modals. To do so, we make use of an experimental design developed in Skovgaard-Olsen et al. (2019) to study individual variation in case of norm conflicts in cognitive psychology, which will be introduced below. But first we introduce the semantic opposition between the three views. Next, we reanalyze previous published findings on epistemic modals and show that they are compatible with individual variation in participants' interpretations. Finally, we use these results to motivate our empirical studies.

Three Opposing Views

The basic problem can easily be appreciated. Different people have access to different evidence and so know, are uncertain, and ignorant of different things. So different people will be prone to describe one and the same situation using different – indeed apparently contradictory – epistemic modals. Consider a simple example. Bill and Doris have arrived in an art museum with three big exhibition halls, A, B, and C. They want to see the new Picasso, which happens to be in hall C, though Bill and Doris do not know this. Since their state of knowledge does not exclude any of the three possibilities, it is epistemically possible for them that the new Picasso is in hall A, possible that it is in hall B, and possible that it is in hall C.

Bill says to Doris:

- (1) The Picasso might be in hall A. *(warranted assertion)*

Meanwhile, Anne and Charlie are also at the museum (Figure 1). Suppose that Anne and Charlie do not know Bill and Doris and are not part of their conversation, but Anne and Charlie can still hear what they are saying. They have already searched room A for the new Picasso without finding it, and Anne, somewhat flustered, whispers to Charlie:

- (2) The Picasso can't be in hall A; it must be in either B or C. *(counterclaim)*

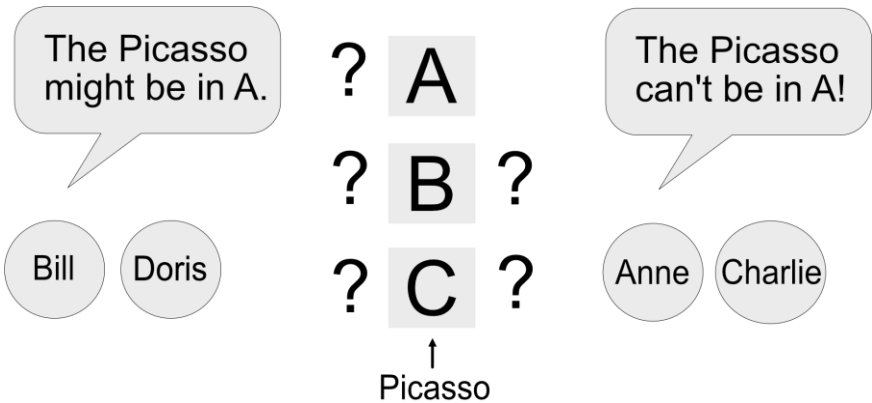


Figure 1. Picasso example.

On the surface, (1) seems to contradict (2); for, at the very least, the claim that the Picasso both might and can't be in hall A is a contradiction. If then Bill's and Anne's claims are jointly contradictory, then one of the claims is false. Now, if one of them is wrong it is presumably Bill; for he said that the Picasso might be in hall A even though it is in hall C (as we know but he doesn't). The problem is that Bill's claim seems perfectly warranted. He knows that the museum has a Picasso but not where. Of course, given what he knows, Bill is not in a position to assert that the Picasso is in hall A, that would be an epistemic lapse on his part. But on any account, knowing nothing more than that the Picasso is in one of the halls, Bill should accept that the Picasso might be in hall A, just as he should accept that it might be in B or C. So, Bill seems to be doing nothing wrong – indeed seems to be doing exactly what he should be doing – yet allegedly by doing so asserts a falsity. That is puzzling; for it is odd that one can reach a falsity by drawing the conclusion that one should draw based on correct (though incomplete) information.

There are a variety of semantic theories that attempt to explain the apparent mystery of might-modals. In our experiments, we focus on Contextualism, Relativism, and Objectivism, which we introduce below. These theories have hybrid versions, which complicate the picture, but here focus on their most simple versions to emphasize their diagnostic features.

Following a standard approach in linguistics (Heim & Kratzer, 1998; Portner, 2009), these views are explicated by formulating truth conditions: states of affairs that make might-statements uttered at a context true or false. These semantic theories were introduced to capture specific norms of language use, which end up having widely different implications for argumentation with epistemic modals, as we shall see. In Appendix 1, the formal details of these views are reviewed, and we more carefully state how subtle differences in the semantic definitions are related to conflicting norms of language use.

Contextualism

Until recently the dominant explanatory framework in philosophy and linguistics was Contextualism (Hacking, 1967; Teller, 1972; Kratzer, 1977; DeRose, 1991). According to Contextualism, the truth conditions for assertions of sentences like (1) and (2) depend on what is known in the *context of use*. Accordingly, the content of an epistemically modal sentence varies with what is known in the context in which it is used. In Bill and Doris' context, his claim amounts to the claim that for all Bill and Doris know, the Picasso is in hall A. In Anne and Charlie's context, her assertion amounts to the claim that she knows that the Picasso is not in hall A. These two claims do not contradict each other. Indeed, they are both true, they have just been asserted in different contexts with different states of knowledge. The puzzle has been removed.

Note, by way of contrast, what would happen if the two couples had met, and Bill, addressing Anne, had asserted (1). It would seem perfectly legitimate for Anne to reply:

(3) No, not true, the Picasso can't be in hall A. We searched (*disagreement in*
hall A thoroughly and it isn't there, it must be in either B or C. *joint context*)

Anne's seemingly reasonable response seems to be a flat denial of Bill's claim; so, it no longer makes sense to treat Bill's claim as a claim about what he and Doris knows. This can be explained by the shift in context. According to Contextualism, once Anne and Charlie have joined the context, Bill's assertion amounts to the claim that for all they (Bill, Doris, Anne and Charlie) jointly know, the Picasso is in hall A, and as Anne and Charlie know that it is not in hall A, Bill's claim is just not true. Whereas (1) and (2) did not contradict each other when asserted in different contexts, they do contradict each other when asserted in the same context.

By allowing flexibility in what counts as known in a context of use, Contextualism obtains considerable degrees of freedom for explaining both what can be plausibly asserted in a given context and how such assertions can be assessed by the audience in the context. Of

course, critics are liable to see this flexibility as a weakness, since it can be interpreted as vagueness concerning what demarcates the context of use and how it changes (Egan & Weatherson, 2011; MacFarlane, 2014). Yet, even if we permit this flexibility, there remain cases that are hard for Contextualism to explain. For consider if Bill, having learned that Anne and Charlie are also looking for the Picasso, said the following:

(4) Have you checked hall A? The Picasso might be there. *(assertion without common knowledge)*

By his initial question Bill plausibly makes it clear that he does not exclude the possibility that Anne and Charlie have checked hall A and might know that the Picasso isn't there. So, Bill clearly doesn't intend his might-statement to be a claim about what they all jointly know. However, Anne's response (3) that flatly denies Bill's might-statement still seems to be perfectly legitimate. This presents a problem for Contextualism. If "The Picasso might be there" in Bill's mouth means that no one in the present company has excluded the possibility that the Picasso is in hall A, then Bill has no business in making the claim. However, if it instead means that he has not excluded the possibility that the Picasso is in hall A, then Anne has no business contradicting him; for he is merely making a claim about himself. Either way Contextualism has problems finding a contextually determined body of knowledge that at the same time makes both the initial assertion and the subsequent assessment plausible. Perceived problems with the contextualist account have opened for alternatives,² which is now an active research area in formal semantics and linguistics (see e.g., Egan & Weatherson, 2011) that is ripe for empirical investigation. Since Contextualism treats 'might-*p*' as a claim to the effect that *p* is consistent with what is known by a contextually determined group of people, for-all-we-know statements will feature as a baseline in our experiments below. Of the three

² There are, however, extensions of the contextualist paradigm that seek to address these issues in a broadly contextualist way. See, for instance, Kratzer's discussion in (2012, p. 100) and von Stechow and Gillies (2008, 2011).

semantic theories investigated, it is a distinctive prediction of Contextualism to treat ‘might’ and ‘for-all-we-know’ statements as equivalent.

Relativism

One influential alternative is assessment Relativism, or Relativism for short (Egan, Hawthorne, & Weatherson, 2005; Kölbel, 2003, 2015a, 2015b; Egan, 2007; Weatherson, 2009; MacFarlane 2011, 2014). While Contextualism holds that the content of a might claim depends on what is known in the *context of use* (c_1), Relativism denies this. Instead, the relativist holds that the truth value of a might-claim depends on what the person assessing the claim knows: truth must be relativised to the *context of assessment* (c_2), shifting the emphasis from the speaker to the ‘listener’ (assessor). For the speaker, of course, the context of use and the context of assessment is the same. So, when Bill asserts (1) his claim is true-for-him, but the same claim is false-for-Anne. Hence, Bill can properly make his claim, and Anne can properly deny it. Importantly, for this brand of relativism, it is not the content that is relative; both Bill and Anne are taking up attitudes towards the same content. Rather, the proposition expressed by the claim uttered at c_1 is thought to have a truth value that is relative to the context of assessment, c_2 .

There are further examples where Relativism seems to do better than Contextualism in accounting for certain kinds of intuitions regarding plausible exchanges involving epistemic modals. Well-known examples are cases of *eavesdropping* and, importantly, *retractions*. Both are discussed further below and will feature centrally in our experiments.

To specify the norm of retraction, MacFarlane (2014, p. 108) distinguishes between the context at which the original assertion was made, c_1 , and the context at which the retraction takes place, c_2 . The norm is then formulated as follows (where p is a placeholder for statements like "The Picasso might be in hall A").

Retraction Rule. An Agent in context c_2 is required to retract an (unretracted) assertion of p made at c_1 if p is not true as used at c_1 and assessed from c_2 .

Retracting is here not the same as admitting fault since the assertion may have been reasonable in the context in which it was made (c_1). Yet, since the assertion is not true as assessed from c_2 its conversational effects need to be undone, according to Relativism.

To illustrate retractions: Upon hearing Anne's reply (3) in the enlarged context to Bill's initial assertion (1), Bill will realize that the Picasso is not in hall A. He should then be willing to retract his earlier claim by conceding something like:

(5) Oh, then I guess I was wrong. *(retraction)*

For the relativist such a retraction makes sense, because while (1) was true-for-Bill when he made his assertion, it is no longer true-for-Bill after Bill has learned that the Picasso is not in hall A. By contrast, the contextualist faces difficulties explaining why such a retraction would be reasonable: Bill's initial claim was a true claim about what he knew at the time, so why retract? Relativism is attractive to the extent that it is better than Contextualism at explaining our intuitions in these cases. Whether these intuitions are shared by ordinary people will be examined in our experiments below.

However, relativising truth to a context of assessment opens a bundle of foundational problems having to do with the regulatory or normative role of truth in a discourse. For instance, we posed the problem of might-modals in terms of the apparent conflict involved in holding that Bill seems to be fully justified in believing and asserting (1) even though the resulting claim is false; based only on correct information he draws the conclusion that he should draw, but the conclusion is false. Relativism resolves this issue by letting (1) be true-for-Bill when it is asserted. So, from Bill's perspective he is saying something true. Yet, on the relativist account, it is false-for-Anne. Indeed, given that 'we' (the readers and writers of this paper) know that the Picasso is not in hall A, what Bill said is also false-for-us.

Accordingly, Relativism introduces relative truth values in accounting for a fragment of natural language, which can shift with changes to information states. Whether this controversial theoretical innovation can be substantiated empirically remains to be seen.

Objectivism

Cantwell (forthcoming) has suggested an analysis that embraces the puzzle: Bill is fully justified, but wrong. With the idea that there is nothing inherently wrong about having a false might-belief or asserting a false might-modal. This might seem contradictory at first, but ‘being wrong’ is ambiguous in this context. It can mean that one has a belief or made an assertion that one shouldn’t have, or it can simply mean that one has a belief or made an assertion that is false. The former is a normative claim, the latter is a purely descriptive claim about the semantic status of a belief or assertion.

By contrast, in normal factual discourse the normative and descriptive uses of ‘being wrong’ seldom come apart; for although false factual beliefs are often excusable, we expect people who realise that they have formed a false belief to make some adjustment to their belief-forming habits; at the very least not to make the same mistake again. However, with might-modals, things are different: the next time Bill finds himself uncertain about which of various scenarios obtain we expect him to conclude that each scenario might obtain, which, in effect, is to make the same ‘mistake’ again. If it is indeed not a mistake, then drawing an erroneous might-conclusion is different from making a factual mistake.

In Cantwell (forthcoming) this way of understanding epistemic modals is dubbed Objectivism. The account holds that there is an objective sense in which someone like Bill is fully justified in his might-judgment yet allows that there is an objective sense in which might claims have a truth value which does not vary with either the context of use or the context of assessment. Instead, its truth or falsity is determined by what someone at the end of enquiry who knew all the facts would judge true. This ideal information state at the end of enquiry

explains how different agents can converge in their modal knowledge. However, Objectivism draws a sharp distinction between the objective property that our modal beliefs are true or false and norms concerning the correct use of might statements.

Applied to the present example: as the Picasso is in hall C, there is an objective sense in which it can't be in hall A, and in this sense it is false that it might be in hall A. On this 'objectivist' analysis, Bill is fully justified in asserting either (1) or (4), and Anne is fully justified in responding with (3). Upon hearing Anne's response (and so learning that the Picasso is not in hall A), Bill is in a position to judge that his previous assertion was false. To the extent to which (5) expresses this, Bill is in a position to assert (5). Yet, this does not amount to an admission of any wrongdoing. To make sense of this situation, Objectivism introduces the notion of *faultless, false* modal beliefs in the case of a previous faultless state of ignorance, and sharply distinguishes between its semantic properties and epistemic evaluation. (See Appendix 1 for further details.)

Reanalyzing Published Findings

Previously published findings on the opposition between Contextualism and Relativism gives a rather unclear picture with results not clearly favoring either theory. We will illustrate this point by considering the studies of Knobe and Yalcin (2014) and Khoo and Phillips (2018). Since the authors helpfully made their original data sets available, we will briefly reanalyze their results with an eye to individual variation.

The papers perform statistical comparisons based on aggregated statistics like the following:

Table 2. Summary Statistics of Previously Published Findings

Condition	Knobe & Yalcin (2014)			Khoo & Phillips (2018)	
	Exp2	Exp3	Exp3	Assessment	Utterance
Factual	<i>True</i> M = 2.03, SD = 1.83	<i>False</i> M = 6.77, SD = 0.62	<i>Retract</i> M = 6.53, SD = 0.94	M = 5.89, SD = 1.39	M = 6.03, SD = 1.27
Modal	M = 4.86, SD = 2.34	M = 3.19, SD = 2.31	M = 4.04, SD = 1.63	M = 4.65, SD = 2.14	M = 4.13, SD = 2.06
Indexical	//		//	M = 5.67, SD = 1.40	M = 2.64, SD = 1.97
DV	“statement is true/false?”		“should retract?”	“At least one of the claims must be false?”	
Scale	7-point Likert Scale from 1 (“completely disagree”) to 7 (“completely agree”)				

Note. In Knobe and Yalcin (2014): ‘Modal’ = might. In Khoo and Phillips (2018): ‘Modal’ = could; ‘Assessment’ = two speakers make conflicting assessments of assertion by third party; ‘Utterance’ = two speakers make conflicting utterances. ‘DV’ = dependent variable.

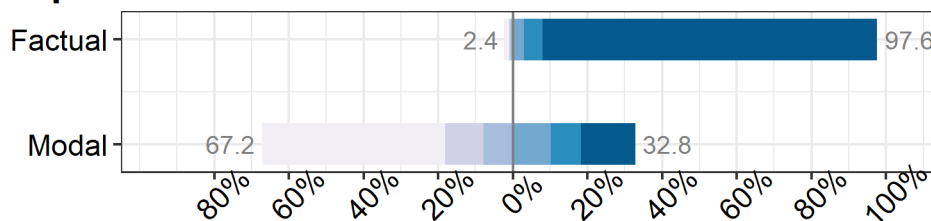
Table 2 displays data concerning whether the statement is true or false, whether the speaker should retract it after a change in information, and whether one of two opposing statements must be false on a Likert-scale from 1-7. In these studies, modal statements (might, could) are contrasted with factual statements and indexical statements (e.g., “I have had breakfast”) to obtain two contrasting baselines. For factual statements, the expectation in Knobe and Yalcin (2014) is that the statement is false and that it should be retracted, and the expectation in Khoo and Phillips (2018) is that participants will agree that one of two contrary statements must be false. For indexical statements, the expectation is that there is no real conflict between contrary statements, and that participants accordingly should disagree with the assessment that at least one of them must be false.

When looking at these data, it is striking that the means for the epistemic modals are at the midpoint of the scales and that epistemic modals have the largest standard deviations. In contrast, the means of the baseline categories are located at more extreme points on the 7 point Likert-scale. Both papers analyzed the results using statistics at the group level. Khoo and Phillips (2018) used a combination of ANOVA and t-tests; Knobe and Yalcin (2014) used ordinal regression. Refitting ordinal regression models with the factors Condition \times Sentence, using the statistical programming language R (R Core Team, 2015) and the R package *brms*

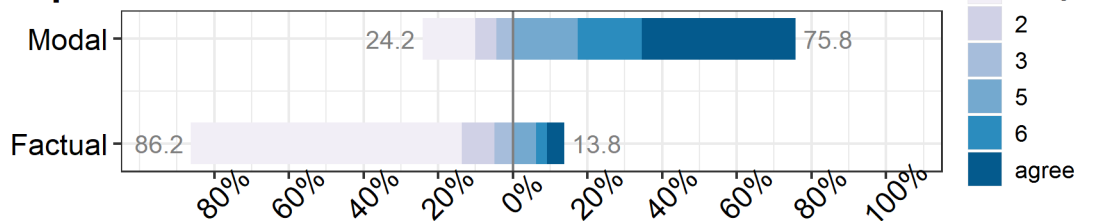
for Bayesian statistics (Bürkner, 2017), yielded the posterior predictive predictions shown in Figure 2.³ Figure 2 confirms that there was a large degree of individual variation in these results for sentences containing epistemic modals, specifically. These patterns of individual variation were not examined in the respective papers, however.

Knobe & Yalcin (2014)

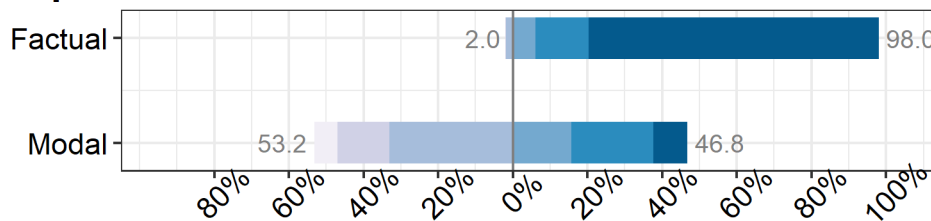
Exp2: False



Exp2: True



Exp3: Retraction



³ See the osf project page for R-scripts: <https://osf.io/g6vym/>.

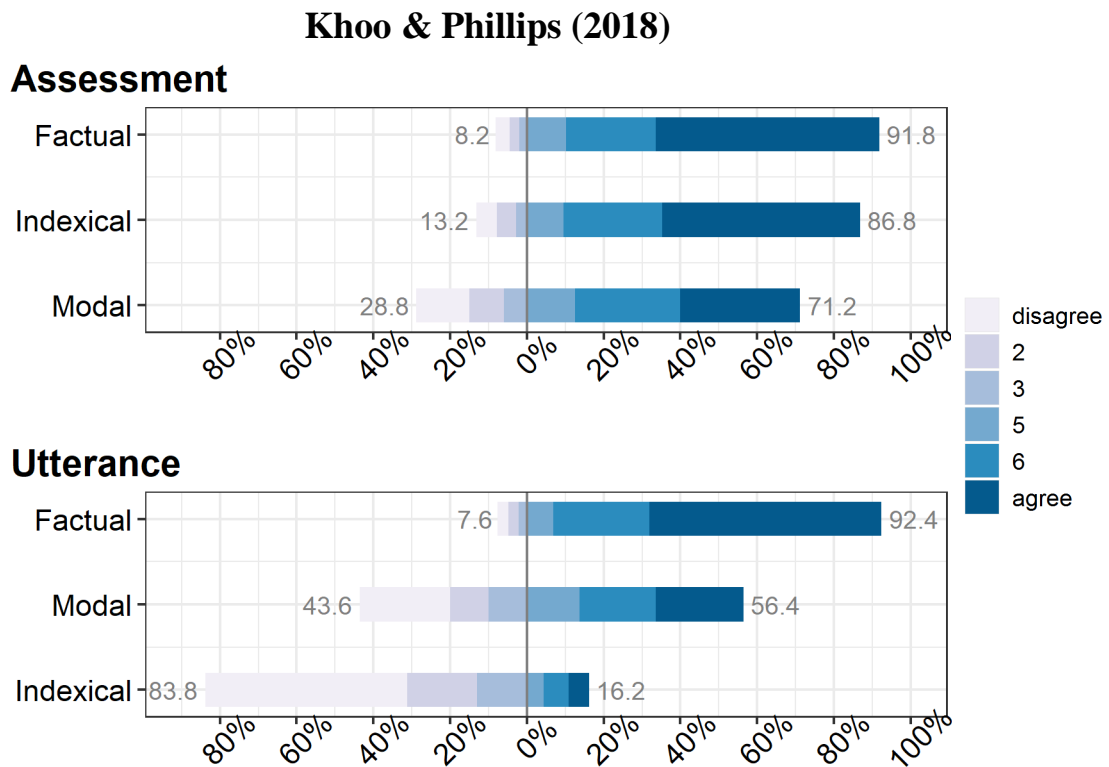


Figure 2. Sampling from the Posterior Predictive Distributions. Predictions based on 500 random sam-ples from the posterior distributions, given the participants did not select the neutral '4' category. For Knobe & Yalcin (2014, Exp2): Condition = True vs. False; Sentence = Modal vs. Factual. For Khoo & Phillips (2018): Condition = Assessment vs. Utterance; Sentence = Factual vs. Modal vs. Indexical. The percentages on the x-axis represent posterior probabilities of selecting categorical outcomes (on the percentage scale). The x-axis extends to 100% in both directions, to represent posterior probability of disagreeing (left) and agreeing (right).

In these plots, values from 1-3 indicate disagreement with the presented statements and values from 5-7 represent agreement. As shown in Figure 2, in each case participants were divided between agreeing (5-7) and disagreeing (1-3) with the presented statements for epistemic modals. Applying statistics based on central tendencies to the data of Table 2 at the aggregate level is liable to misrepresent the distributions in the face of such split tendencies of agreeing and disagreeing. One of the purposes of our experiments is therefore to make a more targeted investigation of individual variation in the interpretation of epistemic modals. At issue is the following underlying assumption:

- (U) There is a uniform interpretation of expressions like epistemic modals. Only one of conflicting semantic theories can be descriptively adequate. If semantic theories like

Contextualism, Relativism, and Objectivism are incompatible, then at most one of them can be descriptively correct.

Through our experiments, we will investigate whether (U) is correct as applied to epistemic modals. In doing so, we apply the individual profiling approach and Scorekeeping Task of Skovgaard-Olsen et al. (2019), which we review below. This leads to our first hypothesis.

(H₁) There is individual variation in the interpretation of might-statements.

Both Khoo and Phillips (2018) and Cantwell (forthcoming) consider that a contributing factor to these mixed results is that participants may have had a difficulty separating truth from justification. On all accounts, the assertability of ‘might *p*’ is determined by the evidence available at the context of utterance. For participants conflating truth judgments with whether a statement is ‘acceptable’ or ‘assertible’, the judgment ‘true’ may be produced in apparent agreement with Contextualism, which only takes information accessible in the context of utterance into account. This conflation hypothesis is accordingly a further alternative hypothesis that we will test in the experiments that follow.

(H₂) Participants conflate assessments of truth and justification.

Both Katz and Salerno (2017) and Khoo and Phillips (2018) argue that data on disagreement concerning epistemic modals is more fundamental than data concerning truth evaluations and retraction. Both studies thus use participants’ judgments about whether two contrary statements can both be true as a measure of adherence to Relativism. Neither study found support for the central prediction of Relativism that the statements “might *p*” and “not-*p*” cannot both be true together, and thus constitute genuine cases of disagreement. But, on the other hand, Contextualism was not favoured either.⁴

⁴ It is moreover likely that the Katz and Salerno (2017) study contains a similar degree of individual variation as the two other studies we have looked at above, which was not captured by the logistic regression analyses reported at the group level. In study 2, Katz and

The Present Studies

One goal of Experiment 1 was to directly probe whether evidence could be obtained for relativistic truth-values, which is the central innovation of MacFarlane (2014) to account for subjective language. As emphasized by Wright (2008: 180): “Correctness varies with context of assessment: that claim is the very heart and soul of truth-relativism. So that is what we need to see evidenced in linguistic practice if linguistic practice is to provide evidence for relativism”. At the same time, Wright (*ibid.*) emphasizes that it must be shown that these evaluations of correctness (or truth) dissociate from evaluations of justifications, as we have seen. Hence, evidence of shifts in truth values across different contexts of evaluations, while controlling for the difference between truth and justification, would constitute direct evidence for the relativistic interpretation of epistemic modals. This leads to our third hypothesis.

(H₃) Shifts in truth-value judgments of might-statements occur across different contexts of evaluations, when controlling for the difference between truth and justification.

In Experiment 1, we investigated whether such truth-value shifts occur and made them the basis for our classification of participants as following Relativism. In addition, we investigated the type of disagreement data which Katz and Salerno (2017) and Khoo and Phillips (2018) take to be fundamental to the debate between Contextualism and Relativism.

Given the possibility of individual variation raised by our reanalysis above, a further objective of Experiment 1 was to probe whether the population could consistently be classified as a following different latent classes of diverging interpretations of epistemic modals. This leads to our fourth research hypothesis as a more precise version of (H₁).

Salerno (2017) thus found that participants had mostly a ca. 50% chance of responding ‘1’ on a binary variable across conditions and items. The Katz and Salerno (2017) thus presents further motivation for investigating patterns of individual variation.

(H₄) Participants can be classified as following different latent classes of diverging interpretations of epistemic modals, which differently influence evaluations of truth and disagreement.

Finally, we tested whether participants thus classified were consistently following their assigned interpretation of epistemic modals when examining the type of retraction data emphasized in both MacFarlane (2014) and Knobe and Yalcin (2014) as being central to the opposition between Contextualism and Relativism.

Norm Conflict Experiments and the Problem of Arbitration

The research question into individual variation and latent classes of diverging interpretations of epistemic modals is motivated by a more general problem affecting the use of norms in empirical studies of rationality. Since many tasks have multiple norms that could be applied to them, Elqayam and Evans (2011) argue that the problem of arbitrating between competing norms is one of the main problems in cognitive psychology preventing the application of norms of judgment and decision-making in experimental research. In contrast, the present experiments aim to show that by classifying participants according to different competence profiles based on divergent norms, the possibility of competing norms (like in Table 1) can be utilized for studying individual variation.

Through our experiments, we show how this problem of arbitration can be handled via Bayesian computational modelling and innovations in the experimental design, which were first employed in the context of conditionals and the psychology of reasoning in Skovgaard-Olsen et al. (2019). Table 3 outlines how this approach applies to Experiment 1.

Table 3. Individual-Profiling via Norm Conflicts

Experiment 1	Goal	Method
Session 1 <i>Phase 1</i>	Estimate group-specific posterior probabilities of truth, justification, and disagreement based on latent classes from phase 2.	Bayesian hierarchical latent trait model.
<i>Phase 2</i>	Classify participants into latent classes based on their reflective attitudes elicited by the Scorekeeping Task.	Bayesian latent class analysis.
Session 2 <i>Phase 3</i>	Probe whether participants follow the norm of retraction based on their individual profiles.	Bayesian mixed linear models.

Note. Details on the Bayesian latent class analysis and the hierarchical latent trait model can be found in Appendix 2. See our osf project page for R-scripts of all the analyses: <https://osf.io/g6vym/>.

The Bayesian latent class model is presented below and the hierarchical latent trait model is explained in Appendix 2. On this approach, participants are classified in Session 2 according to their reflective attitudes in situations of norm conflicts. Via latent class analysis, it is probed whether individual profiles of the participants can be established that follow Relativism, Contextualism, and Objectivism, respectively. Their latent classes are used to estimate group-specific means in participants' truth, justification, and disagreement judgments. Participants' reflective attitudes are elicited via the Scorekeeping Task, where participants are put in the position of a scorekeeper. The scorekeeper has to assess the performance of fictive participants, who have produced incompatible responses to the task that the participants have just completed. Through this task, participants commit to an interpretation of epistemic modals by criticizing and sanctioning their fictive peers based on their mutual criticism. A comparison is then made between participants' reflective attitudes with their own case judgments to investigate how well they match.

In session 2, it was probed whether participants consistently followed the assigned profiles to apply the associated norms in a novel task. For the present case of norm conflicts with epistemic modals, this involves investigating whether participants follow the norm of retraction. Participants classified according to Objectivism and Relativism are predicted to enforce this norm, whereas participants following Contextualism are predicted to flout it.

Experiment 1

Session 1

The goal of Session 1 was to analyze participants' judgments of truth and disagreement as influenced by latent classes established in the Scorekeeping task. As we will see, the classifications were mutually exclusive, which means that a participant could at most be assigned to one latent class (Contextualism, Relativism, Objectivism).

Method

Participants

The experiment was conducted over the Internet to obtain a large and demographically diverse sample. A total of 780 people completed the experiment. The participants were sampled through the Internet platform Mechanical Turk from the USA, UK, Canada, and Australia. They were paid a small amount of money for their participation (on average 6\$ per hour) and told that there would be in addition be a 1\$ bonus, if they answered accurately and participated in the second half one week later. The following *a priori* exclusion criteria were used: not having English as native language, completing the task in more than 2 standard deviations below or above the mean completion time, failing to answer at least one of two simple SAT comprehension questions correctly in a warm-up phase, and answering 'not seriously at all' to the question 'How seriously do you take your participation' at the beginning of the study. Since some of these exclusion criteria were overlapping, the final sample for session 1 consisted of 540 participants. Mean age was 39.42 years, ranging from 18 to 78. 48.15 % of the participants self-identified as male, 51.67% self-identified as female, and one person preferred not to identify with either gender. 79.44 % indicated that the highest level of education that they had completed was an undergraduate degree or higher. Applying the exclusion criteria had a minimal effect on the demographic variables.

Design

Phase 1 of the experiment had a within-subject design with two factors: Sentence (with three levels: factual vs. might vs. know) and type of dependent variable, DV (with four levels explained below: justified vs. true_{t1} vs. true_{t2} vs. true_{both}). Since all four dependent variables were presented on every trial, and we wanted four trial replications for each cell of the design, each participant in total went through 48 within-subject conditions.

Procedure Shared by Session 1 and 2

The four trial replications of the three Sentence within-subject conditions were randomly assigned to 12 different scenarios. Random assignment was performed without replacement such that each participant saw a different scenario for each condition. This ensured that the mapping of condition to scenario was counterbalanced across participants preventing confounds of condition and content. To reduce the dropout rate during the experiment, participants first went through three pages stating our academic affiliations, posing two SAT comprehension questions in a warm-up phase, and presenting a seriousness check asking how careful the participants would be in their responses (Reips, 2002). Participants were then asked to supply their Mechanical Worker ID so that their responses in Session 1 could be matched with Session 2, one week later.

Materials and Procedure Specific to Session 1

Phase 1: The Eavesdropper Task

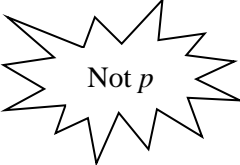
The list of the 12 scenarios used can be found on the *osf* page.⁵ The scenarios used were inspired by the stimulus materials of Knobe and Yalcin (2014) and Katz and Salerno (2017), but further were created to increase the number of items. Since the task involved time indices printed in writing and illustrated on analog clocks, participants first saw three practice items to ensure that they had understood the setup. In addition, participants were debriefed after the practice items to make sure that they paid careful attention to such distinctions as

⁵ https://osf.io/g6vym/?view_only=9425e07e499949b9971baffd1c5519a5

whether a statement (marked in blue) was true before learning about an additional fact and whether the statement was true now, after the additional fact had been learned.

In the first phase, participants were presented with the 12 scenarios in random order, displayed in the format illustrated below. These scenarios concern so-called “eavesdropper” cases, where a person, who is not taking part in a conversation, possesses additional factual information, which can be used to evaluate the assertions of the speakers. In this task, participants are first presented with the information used as the basis for the assertions by the speakers (at t_1), and then there is a continuation, where the factual information of the eavesdropper is added (at t_2), as illustrated in Table 4 below.

Table 4. Temporal Format of the Eavesdropper Task

Statement (S)	Continuation (Fact)
<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;"> <p><i>p / might p / for all we know, p</i></p> </div> <p>14:10 (t_1)</p>	<div style="text-align: center;">  <p>Not <i>p</i></p> </div> <p>14:12 (t_2)</p>
<p><i>Is the statement justified?</i></p>	<p><i>Was the statement true at 14:10? Is the statement true now at 14:12? Is it possible for both the statement and the continuation to be true at the same time?</i></p>

This temporal format was used to manipulate changes in information states, which figures centrally in the semantic opposition between Contextualism, Relativism, and Objectivism (Egan & Weatherson, 2011).

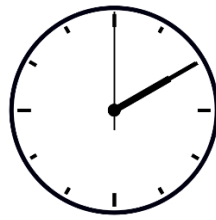
To emphasize certain details of the scenario, some phrases were highlighted in blue and red (as indicated below). Here is an example of a scenario:

Heth and Jed are at the coffee shop together waiting for Fred, who will be reciting poetry later that night. A man is approaching, but Jed has poor eyesight and can only see the bare outlines of the man and says:

14:10: “That person **MIGHT** be Fred.”

[/"FOR ALL WE KNOW, that person is Fred"]

[/"That person is Fred"]



Participants were then asked four questions in the following order. First, they were asked whether they agreed/disagreed with the statement that “The speaker is justified in making the blue assertion”. Next, the time displayed on the analog clock shifted by two minutes, and they were presented with the following continuation of the scenario for the remaining three questions:

Continuation (2 min later):

Cindy works for security at the coffee shop and is watching the room by video camera from a remote location. Her camera has a close-up of the approaching man, and while eavesdropping on Heth and Jed’s conversation Cindy says to herself:

14:12: “That person can’t be Fred. I know Fred has a scar on his face.”

Participants were then asked to indicate whether they agreed/disagreed with the following statements after being presented with the continuation:

“**Before** learning about the continuation, the blue statement **was** true at 14:10”.

“**After** learning about the continuation, the blue statement made at 14:10 **is** true **now** at 14:12.”




“It is possible in this case for both the claim marked in blue and the statement made in the continuation to be true (at the same time).”

Each of the four DVs was presented as indicator variables (“yes” vs. “no”) to the participants.

Phase 2: The Scorekeeping Task

After completing this task for the 12 different scenarios, participants were presented with the Scorekeeping Task following Skovgaard-Olsen et al. (2019). In the Scorekeeping Task, participants were told that when given the task that they had just completed, John, Robert, and Simon responded very differently. Participants were then presented with one of the scenarios that they had responded to in the Might condition. After having seen the first part of the scenario *before* the continuation, participants could click “Continue” to see the next part *after* the continuation. By further “Continue” clicks, participants could elicit the three responses by John, Robert, and Simon in random order, as shown in Table 5.

Table 5. The Scorekeeping Task, the Three Conflicting Responses

	<p>John responded:</p> <p>At 14:10 the blue statement was true. After learning about the continuation at 14:12 the blue statement is no longer true.</p>
	<p>Robert responded:</p> <p>At 14:10 the blue statement was true. After learning about the continuation at 14:12 the blue statement is still true.</p>
	<p>Simon responded:</p> <p>At 14:10 the blue statement was not true. After learning about the continuation at 14:12 the blue statement is still not true.</p>

Note. The response of John represents Relativism, the response of Robert represents Contextualism, and the response of Simon represents Objectivism. As above, various phrases were highlighted in red and blue to make the processing easier for participants.

Participants were then instructed that “Robert, John, and Simon cannot all be right!” and that they would see each of their responses repeated along with a criticism by the two other persons. Participants were told that their task was to decide based on these mutual criticisms whose response was the most adequate and whose “HIT” should be approved.







On Mechanical Turk, tasks are described as ‘HITs’ (Human Intelligence Tasks) to participants. Since participants are financially rewarded by approvals of HITs, and build up a reputation on this basis, the approval of HITs was used as an ecologically valid sanctioning

measure in Skovgaard-Olsen et al. (2019) that participants are motivated to reason about.

Here we adopt this measure as well.

The three criticisms were presented in random order on three consecutive pages. On each page, participants were first presented with the scenario and the response being criticized for repetition. Next, they were presented with the corresponding criticism of the response, by the two other parties, as illustrated in Table 6 below, and asked whether they found the given criticism compelling (“yes” vs. “no”). As above, each bit of information on the page was elicited by the participant by pressing “Continue”. Screenshots of the pages can be obtained at the *osf* project page.⁶

Table 6. The Scorekeeping Task, the Mutual Criticisms

	Robert and Simon’s criticism of John:
	“How can you both say that the blue statement was true at 14:10 and is false at 14:12? That makes no sense!”
	John and Simon’s criticism of Robert:
	“How can you say that the blue statement is true at 14:12 after the continuation has become known? That makes no sense!”
	John and Robert’s criticism of Simon:
	“How can you say that the blue statement was false at 14:10 before the continuation became known? That makes no sense!”

Note. The response of John represents Relativism, the response of Robert represents Contextualism, and the response of Simon represents Objectivism. As above, various phrases were highlighted in red and blue to make the processing easier for the participants.

Finally, participants were presented with a page in which all three responses were repeated with the instruction that they should indicate whose “HIT” on Mechanical Turk should be approved after having seen John, Robert, and Simon’s mutual criticism. Out of the following

⁶ https://osf.io/g6vym/?view_only=9425e07e499949b9971baffd1c5519a5

list displayed in random order, participants could select one option for approval: 1) Simon’s HIT, 2) John’s HIT, 3) Robert’s HIT. Participants were then asked a few demographic questions.

Results and Discussion

Phase 1 Judgments

Table 7 displays some initial descriptive statistics, which report the central tendencies of the four dependent variables across the three types of sentences at the aggregate level.

Table 7. Descriptive Statistics of Phase 1

	Justified	True at t₁	True at t₂	Both True
Factual	M = 0.75 (0.43)	M = 0.39 (0.49)	M = 0.12 (0.33)	M = 0.17 (0.37)
Might	M = 0.91 (0.28)	M = 0.71 (0.45)	M = 0.21 (0.4)	M = 0.34 (0.47)
For-all-we-know	M = 0.89 (0.31)	M = 0.71 (0.45)	M = 0.27 (0.44)	M = 0.41 (0.49)

Note. Standard deviations are reported in parentheses.

To investigate whether these data conceal individual variation (H₁), the analysis below models the data as being produced by three latent classes (H₄).

To perform the analysis, we first applied a Bayesian latent class analysis to participants’ responses in the scorekeeping task (see Appendix 2) and then we analyzed the 12 dependent variables in Table 7 based on these three latent classes. Table 8 provides an overview of all these dependent variables and outlines the contrasting predictions of Contextualism, Relativism, and Objectivism based on these measures.

Table 8. Classification and Predictions

Statement (S) made at t ₁ :	Factual <i>p</i>			Might <i>p</i>			For all we know <i>p</i>		
	C	R	O	C	R	O	C	R	O
DV_{True Before} : Was S true at t ₁ ?	0	0	0	1	1	0	1	1	1
DV_{True After} : Is S true after learning that ¬ <i>p</i> at t ₂ ?	0	0	0	1	0	0	1	1	1
DV_{Justified} : Was the speaker justified in asserting S at t ₁ ?	1	1	1	1	1	1	1	1	1
DV_{Both True} : Can both S and ¬ <i>p</i> be true at t ₂ ?	0	0	0	1	0	0	1	1	1

Note: $S \in \{\text{factual } p, \text{ might } p, \text{ for-all-we-know } p\}$ ‘C’ = Contextualism; ‘R’ = Relativism; ‘O’ = Objectivism. ‘ t_1 ’ = time of utterance of S. ‘ t_2 ’ = time of evaluation of evaluation of the statement, S, after learning that $\neg p$.

Since there were four dependent variables for each type of Sentence (might p , for-all-we-know p , and factual p), 12 binominal rate parameters were estimated for a given participant based on four trial replications with unique scenarios.

In Knobe and Yalcin (2014), the focus was on truth value judgments at t_2 (in addition to retraction judgments, which we will return to in Session 2). In Khoo and Phillips (2018), the focus was on incompatibility judgments at t_2 . Table 8 integrates both these judgments and additionally makes a classification based on possible shifts in truth value judgments by measuring truth values at both t_1 and t_2 to test (H₃).

Furthermore, Table 8 controls for the possibility of conflating truth and justification by measuring both to probe (H₂), the conflation hypothesis (Khoo & Phillips 2018; Cantwell, forthcoming). Finally, Table 8 includes a subjective (“for all we know p ”) and an objective (“factual p ”) baseline to compare might-statements with. As can be seen from Table 8, the predictions for these baselines remain invariant across the three types of interpretations of might-statements. We here follow the semantic treatment of “for all α knows p ” given in MacFarlane (2014, p. 265), according to which the statement is true iff p is not excluded by what α knows at t_1 at the context of utterance. According to this definition, further information acquired at t_2 in the eavesdropper cases should not have an impact of the truth value of for-all-we-know statements uttered at t_1 , since the statement restricts its scope to what was known back then. On this basis, the following hypothesis can be formulated.

(H₅) Only participants classified as following Contextualism treat might-statements and for-all-we-know statements alike.

Moreover, the justification of might-statements is also non-diagnostic w.r.t. these three interpretations. Rather, the distinguishing features concern: 1) whether both the might-statement (S) and the revealed fact in the continuation ($\neg p$) can both be true at the same time,

and 2) whether the might-statement is true at the two time points (t_1 before $\neg p$ was revealed, and t_2 after $\neg p$ was revealed). Only Contextualism permits might-statements and $\neg p$ to be true at the same time. Only Contextualism allows might-statements to remain true at t_2 after $\neg p$ has been revealed. In contrast, Relativism posits as its distinguishing feature that a shift in truth values of might-statements occurs between t_1 and t_2 . The distinguishing feature of Objectivism is to deny that the might-statements were true at t_1 before $\neg p$ was revealed. These predictions permit us to test (H₄) that participants can be characterized by following different latent classes in their interpretations of might-statements.

Phase 2: The Scorekeeping Task

Table 9 displays descriptive statistics of the central tendencies at the aggregate level.

Table 9. Descriptive Statistics for Phase 2

	Criticism of	HIT approval of
Relativism	M = 0.30 (0.46)	M = 0.58 (0.49)
Contextualism	M = 0.66 (0.47)	M = 0.23 (0.42)
Objectivism	M = 0.67 (0.47)	M = 0.18 (0.39)

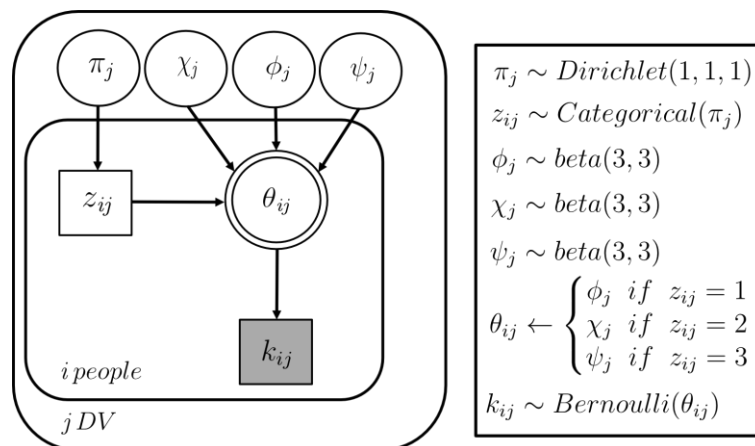
Note. Standard deviations are reported in parentheses.

To investigate whether these data conceal individual variation, the analysis below applies a latent class analysis to identify three latent classes.

Latent Class Analysis

A latent class analysis (Mair, 2018, Li et al., 2018) was applied to participants' responses in the Scorekeeping task. The latent class analysis was implemented in a Bayesian framework was fitted in a Bayesian framework through a Gibbs sampler via JAGS (Plummer, 2019), which estimates the posterior distributions of model parameters by means of Monte Carlo-Markov chains. The latent class analysis in Table 10 was applied to the 6 scorekeeping judgments of the participants.

Table 10. Latent Class Analysis



Note. Table 10 displays the latent class analysis with 3 groups. In addition, versions with 1, 2, and 4 groups solutions were fitted to the $j = 1, \dots, 6$ scorekeeping judgments in a model comparison.

This latent class analysis was repeated for 1, ..., 4 group solutions and a comparison in model fit of the resulting models was made (Table 11). Table 11 reports two information criteria. Of the two, Vehtari et al. (2017) recommend relying on LOOIC which is based on leave-one-out-cross validation. Applying this information criterion, the model comparison in Table 11 showed a clear preference for a three class solution.

Table 11. Latent Class Models

	WAIC	LOOIC	Δelpd (SE)	Weight
LCA ₁	3868.3	3868.3	-968.29 (26.38)	0
LCA ₂	2611.5	2678.7	-373.48 (15.67)	0
LCA ₃	1919.2	1931.7	--	1.00
LCA ₄	1922.2	1970.9	-19.60 (1.34)	0

Notes. LOOIC = leave-one-out cross-validation information criterion. WAIC = Watanabe-Akaike information criterion. ‘elpd’ = expected log predictive density is a measure of the expected out-of-sample predictive accuracy. Note that information criteria can take both positive and negative values and that the lowest value on the real line still indicates best fit. The weights are Bayesian stacking weights based on LOOIC.

A solution with three latent classes was then fitted to the data and posterior item response probabilities for the different items in the Scorekeeping Task were estimated for each of the latent classes along with the frequency of the classes in the population.

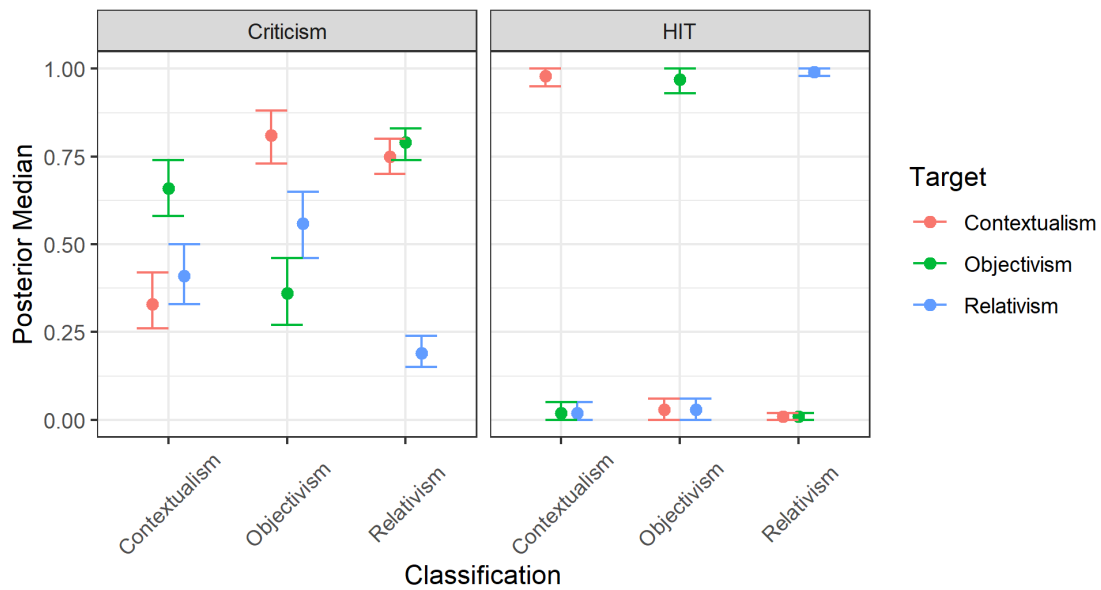


Figure 3. Phase 2 Classification. Posterior Median estimates based on the three latent classes in the Scorekeeping Task: 316 (58%) participants were classified as following Relativism, 128 as following Contextualism (24%), and 96 (18%) as following Objectivism. The parameter estimates indicate the posterior median probabilities of a) finding the criticism of the respective view compelling (yes vs. no coded as 1 vs. 0), and b) approving the HIT of the respective view after having seen the mutual criticism of all sides. The target labels in the legend indicate which semantic interpretation is criticized and approved, respectively. The error-bars indicate 95% highest density interval (HDI).

The latent class identification in Figure 3 was based on the posterior probabilities of both agreeing with the criticism of Relativism, Objectivism, and Contextualism, respectively, and the corresponding HIT approvals (after learning the criticism of all sides). As shown, the posterior median HIT approvals in the phase 2 classification were ca. 99% for all latent classes, and the posterior median probability of accepting criticism of the assigned view was 19%, 33%, and 36% for Relativism, Contextualism, and Objectivism, respectively. This suggests that the scorekeeping data can be used to determine the class membership of each participant. Since the latent class analysis is used to the optimal number of classes that differ on the scorekeeping variables, the differences reported in Figure 3 serve as a sanity check on the separation of the classes. In what follows, we will use these assigned classes to estimate group specific means in participants' phase 1 responses and evaluate contrast effects across classes.

Testing (H₄) by Analysing the Estimated Parameters

The identified latent classes from the scorekeeping task differ on the binominal rate parameters by which participants produce ‘yes’ responses in the four trial replications for each of the four DV in Table 8. To analyze these rate parameters, a hierarchical latent trait model was applied to the data, which used the latent classes from a above as a grouping variable for the latent means (see Appendix 2).

Phase 1 Parameters

The further qualitative differences in Figure 4 indicate that the classifications thus formed are predictive of other judgments of truth and disagreement in support of (H₄).

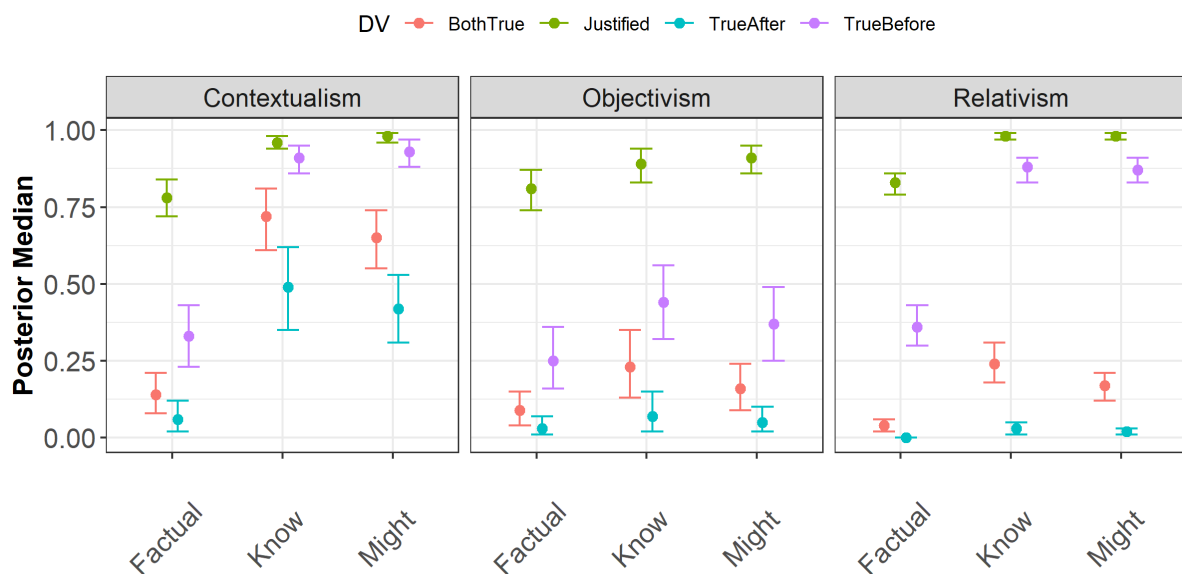


Figure 4. Phase 1 Performance. Parameter estimates of posterior probabilities of answering “yes” for the four DVs in Session 1 for the three latent classes. In total, 540 participants were classified into three groups: Relativism (316), Contextualism (128), and Objectivism (96). ‘DV’ = dependent variable. The error bars indicate the 95% highest density intervals.

Figure 4 shows that the distinctive patterns of Contextualism, Relativism, and Objectivism concerning might-statements outlined in Table 8 are preserved in the estimated posterior median probabilities of answering “yes” for participants classified as belonging to each of these three latent classes.

Model fit was with the posterior-predicted *p* value based on the T₁ posterior model check proposed in Klauer (2010). T₁ measures the adequacy of the models in capturing the

mean observed outcome frequencies (aggregated across persons). A small p value for this test statistics indicates that the posterior predictive distribution of the model fails to capture the aggregate outcome frequencies of the data. It was found that $p_{T1} = .50$, hence the model could well account for aggregate outcome frequencies in the data and it passed this posterior predictive model check.

Next, contrasts effects were calculated for the group-specific means in binomial rate parameters for each of the three latent classes (Table 12). On this analysis, a contrast is interpreted as a credible finding if the 95 % HDI⁷ for the effect excludes 0.

Table 12. Contrasts Effects of Group-Specific Means

Variable	Contrast	Might		For all we know		Factual	
		$\tilde{\Delta}$	95% HDI	$\tilde{\Delta}$	95% HDI	$\tilde{\Delta}$	95% HDI
<i>Both true</i>	$\theta_{Rel} - \theta_{Con}$	-.48	[-.58, -.38]	-.48	[-.60, -.35]	-.11	[-.17, -.04]
	$\theta_{Rel} - \theta_{Obj}$.01	[-.08, .09]	.01	[-.12, .13]	-.06	[-.12, .00]
	$\theta_{Con} - \theta_{Obj}$.49	[.36, .61]	.49	[.32, .62]	.05	[-.04, .13]
<i>Justified</i>	$\theta_{Rel} - \theta_{Con}$.005	[-.01, .02]	.02	[-.01, .04]	.05	[-.02, .11]
	$\theta_{Rel} - \theta_{Obj}$.07	[.03, .13]	.09	[.04, .15]	.02	[-.05, .09]
	$\theta_{Con} - \theta_{Obj}$.07	[.02, .12]	.07	[.02, .14]	-.03	[-.11, .06]
<i>Was true at t_1</i>	$\theta_{Rel} - \theta_{Con}$	-.06	[-.11, .00]	-.03	[-.09, .03]	.03	[-.09, .15]
	$\theta_{Rel} - \theta_{Obj}$.50	[.37, .63]	.44	[.31, .57]	.11	[-.02, .22]
	$\theta_{Con} - \theta_{Obj}$.56	[.43, .69]	.47	[.34, .60]	.07	[-.07, .22]
<i>Is true at t_2</i>	$\theta_{Rel} - \theta_{Con}$	-.40	[-.51, -.30]	-.46	[-.60, -.32]	-.06	[-.12, -.02]
	$\theta_{Rel} - \theta_{Obj}$	-.03	[-.08, .00]	-.04	[-.11, .01]	-.03	[-.07, .00]
	$\theta_{Con} - \theta_{Obj}$.36	[.25, .49]	.42	[.26, .57]	.03	[-.02, .10]

Note. The contrasts were calculated based on the differences in 10,000 random posterior draws of the group-specific mean parameters on the probability scale based on the three latent classes. The posterior medians of these contrast effects are reported together with their 95% highest density interval. ‘Rel’ = Relativism, ‘Con’ = Contextualism, ‘Obj’ = Objectivism.

Focusing on might-statements, the strong signature differences in whether might statements were true before the fact that not- p became known (Relativism - Objectivism, $\tilde{\Delta} = .50$, 95% HDI [.37, .63]; Contextualism - Objectivism, $\tilde{\Delta} = .56$, 95% HDI [.43, .69]) and

⁷ A HDI interval is an interval of the posterior distribution where all points within the interval have a higher probability density than points outside it.

after (Relativism - Contextualism, $\tilde{\Delta} = -.40$, 95% HDI [-.51, -.30]; Contextualism - Objectivism, $\tilde{\Delta} = .36$, 95% HDI [.25, .49]) were found. For Relativism, a stark shift in truth values of might-statements was thus observed in support of (H₃), which spanned most of the probability scale ($\tilde{\Delta} = .85$, 95% HDI [.81, .89]).

Based on the group specific means, the research hypothesis that relativists, contextualists, and objectivists differ on whether two contrary might-statements can both be true was tested. It was found that contextualists had a higher posterior median probability of accepting both of the contrary might-statements as true than relativists (Relativism - Contextualism, $\tilde{\Delta} = -.48$, 95% HDI [-.58, -.38]) and objectivists (Contextualism - Objectivism, $\tilde{\Delta} = .49$, 95% HDI [.36, .61]).

To test the alternative hypothesis that participants conflate truth and justification, we focus on the factual statements as the objective baseline and compare the contrast between $DV_{\text{justified}}$ and $DV_{\text{True Before}}$ across the three latent classes. It was found for Contextualism ($\tilde{\Delta} = .45$, 95% HDI [.35, .55]), Relativism ($\tilde{\Delta} = .47$, 95% HDI [.40, .53]), and Objectivism ($\tilde{\Delta} = .56$, 95% HDI [.44, .65]) that the posterior probability of factual statements being justified before not-p was learned was higher than the probability of factual statements being true. The results thus allow us to reject the hypothesis, (H₂), that participants conflate these two types of evaluation (Khoo & Phillips, 2018; Cantwell, forthcoming).

According to the predictions in Table 8, a distinguishing feature of Contextualism is to treat might-statements and for-all-we-know statements alike – as both expressing contextual features of the present epistemic state. In contrast, it was found across all three latent classes that participants followed the trend of treating for-all-we-know statements in a qualitatively similar way to might-statements, in spite of the stark differences in the interpretations of the latter (Table 12). This finding is surprising and contradicts (H₅) – moreover, it was consistent over several pilot studies. It may indicate that participants could not detect a difference in meaning between might-statements and for-all-we-know statements in a within-subject

comparison, although participants across latent classes displayed sharply differing responses patterns concerning both types of statements.

Session 2

The Norm of Retraction

A week later, the same participants from Session 1 were invited to participate in a task probing how participants would react to violations of the norm of retraction. Session 2 investigated whether participants classified by the latent classes in Session 1 would react to norm violations of the norm of retraction in accordance with the interpretation of epistemic modals assigned. This leads to our sixth hypothesis.

(H₆) Only participants classified according to Objectivism and Relativism enforce the norm of retraction for might-statements.

To investigate whether participants thus classified followed this line of response, Session 2 presented the same participants from Session 1 one week later with a series of items, where the norm of retraction was violated for factual-statements, might-statements, and for-all-we-know statements. It was then investigated how participants reacted to these potential norm violations. The factual statements and the for-all-we-know statements were used as baselines for cases, where the norm of retraction applies or does not apply, respectively, to test for differences between the two and might-statements across the phase 2 classification from Session 1.

In the psychology of reasoning, participants who have received instruction in logic are routinely excluded from participation to tap into participants' natural linguistic competence (see e.g., Evans, 2002; Klauer et al., 2010). In our experiments, we decided instead to include previous exposure to logical training as a covariate in the models to examine empirically whether it played a role for participants' adherence to the norm of retraction. We did this, because we expected that participants, who had received prior training in logic, would be

more likely to exhibit consistent competence profiles when presented with questions that test subtle differences in the truth, justification, and retraction of might-statements than participants, who had not received such training. By including this co-variate in our analysis, we could investigate whether the degree of internal consistency between session 1 and 2 is higher for participants who had received prior logical training (irrespective of their adherence to Relativism, Contextualism, or Objectivism). This leads to our sixth hypothesis.

(H₆) Prior instruction in logic facilitates consistent competence profiles.

Method

Participants

Only participants who had taken part in Session 1 and had not been excluded by the exclusion criteria in Session 1, were invited to participate in Session 2 one week later. Out of the 540 invited participants from Session 1, 461 participants (85.37%) took part in Session 2. The participants were paid on average 6\$ an hour for their participation and an additional bonus of 1\$ for having taken part in both sessions.

Participants who indicated that English was not their native language, that they would not take their participation seriously, or who completed the task in less than 240s or more than 3600s were excluded. The final sample consisted of 438 participants whose responses from both sessions could be identified. Mean age was 39.40 years, ranging from 18 to 78. 47.95% of the participants self-identified as male, 51.83% self-identified as female, and one participant chose not to identify with either gender. 81.05% of the participants indicated that the highest level of education that they had completed was an undergraduate degree or higher. 27.40% indicated that they had previously received prior training in logic.

Design

The experiment had a within-subject design with one factor: Sentence (with three levels: factual vs. might vs. for-all-we-know). To allow for six trial replications for each cell of the design, each participant in total went through 18 trials in random order.

Materials and Procedure Specific to Session 2

The within-subject conditions were randomly assigned to 18 different scenarios (see the *osf* project page⁸ for sample scenarios). The participants first went through two practice trials and were then instructed to pay attention to small details in the sentences highlighted in blue. For each of the 18 scenarios, the participants were first presented with the first part of the scenarios (the content above the text in bold in the example below) and could then click “continue” for the presentation of the rest of the scenario (the content beginning with the text in bold). One of the 18 scenarios was as follows:

During a murder trial the defense lawyer presents evidence that his client did not commit the murder but he is also challenged by the blood found at the crime scene.

At the trial, the defense lawyer says:

“FOR ALL WE KNOW, the blood was planted on the scene of crime”

[/”The blood MIGHT be planted on the scene of crime”]

[/”The blood WAS planted on the scene of crime”]

The prosecutor has the forensic lab conduct a test that can rule out that the blood was planted on the scene of the crime.

In court, the defense lawyer reacts to the forensic report by saying:

Oh really? I guess then the prosecutor is right. But I still maintain it was true that:

“FOR ALL WE KNOW, the blood was planted on the scene of crime”

[/”The blood MIGHT be planted on the scene of crime”]

[/”The blood WAS planted on the scene of crime”]

⁸ https://osf.io/g6vym/?view_only=9425e07e499949b9971baffd1c5519a5

when I said it. I stand by my claim and refuse to take it back.

The second occurrence of the speaker's claim was highlighted in blue, and the task of the participants was to indicate whether they agreed/disagreed ("yes, I agree" vs. "no, I disagree") with the following statement: "The defense lawyer should have taken back his claim by saying 'Oh, then I guess I was wrong'".

The participants completed the 18 within-subject conditions in random order with a unique assignment of the 18 scenarios to each condition. After completing the task, participants were asked a few demographic questions.

Results and Discussion

Session 2 served the goal to measure whether participants react to violations of the norm of retraction for might-statements as a function of their phase 2 classification (H_6), and whether they had received previous logical training (H_7). Accordingly, the main dependent variable was a binary variable ("yes" vs. "no") indicating whether the speaker had violated the norm of retraction by refusing to take back their assertion after the fact $\neg p$ was revealed.

Using the statistical programming language R (R Core Team, 2015), mixed linear models with a binomial likelihood function were applied to participants' responses of violations of the retraction norm across the three types of sentences (factual vs. might vs. for-all-we-know). This analysis was conducted using the R-package *brms* for mixed-effects models in Bayesian statistics (Bürkner, 2017). A model with random intercepts for participants and items was fitted for the phase 2 classification from Session 1. This model permitted the phase 2 classification to interact with the type of sentence evaluated (factual vs. might vs. for-all-we-know) and included participants' logical training (yes vs. no) as a covariate along with its interactions with the Sentence factor.

To get an overview over the effects, we can inspect whether the 95% HDI intervals of contrasts in the parameter estimates include zero (marked by the dashed line in Figure 5).

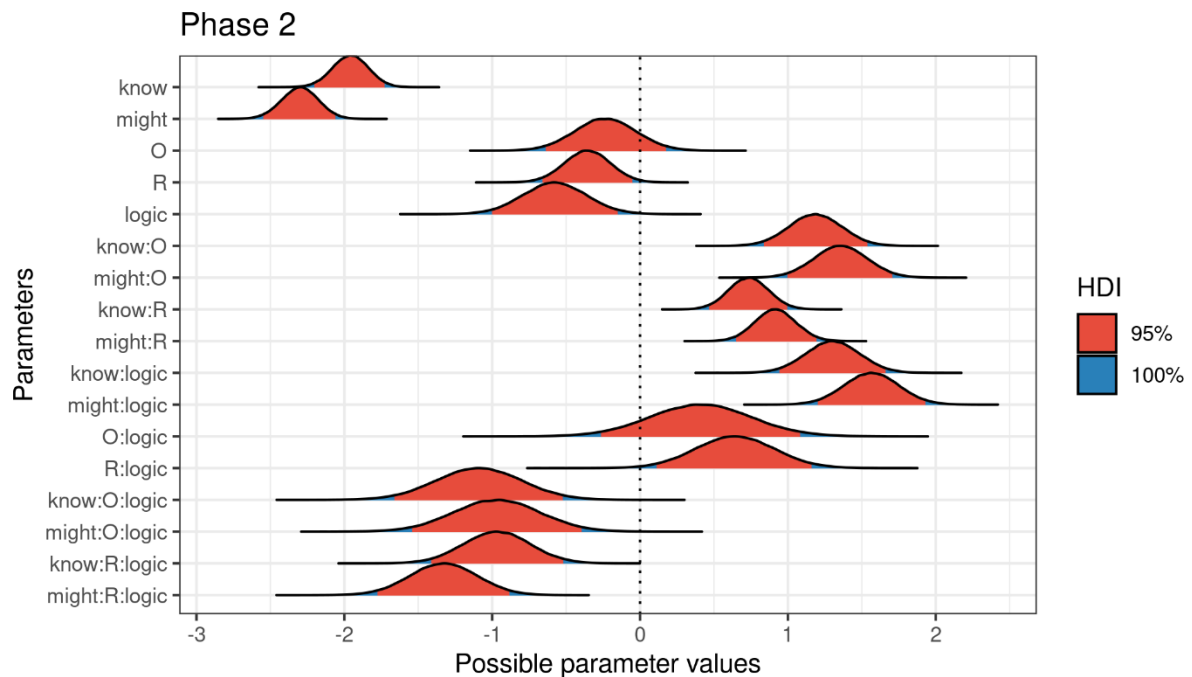


Figure 5. Parameter Estimates and 95% HDI intervals. The plot displays the contrast effects for phase 2. Reference levels: Factual sentences, Contextualism, and no prior training in logic. ‘C’ = Contextualism, ‘O’ = Objectivism, ‘R’ = Relativism.

As the plot in Figure 5 shows, there were credible effects of three and two-ways interaction effects between training in logic, the phase 2 classifications, and the three types of sentences. We will therefore illustrate the effects by presenting a plot of the posterior predictions (Figure 6) as a visual aid.

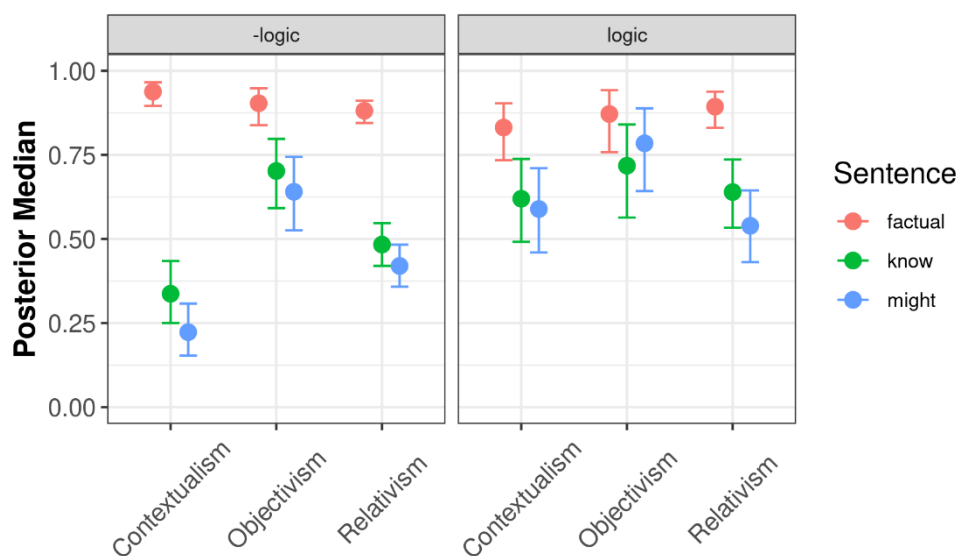


Figure 6. Session 2: Norm of Retraction. Posterior medians of the probability of objecting to violations of the norm of retraction for factual statements, for-all-we-know statements, and might-statements across the phase 2 classification from session 1. The left-side displays the

posterior predictions for participants who had not previously received instruction in logic; the right displays the predictions for those who had. The error-bars display 95% credible intervals.

Across the phase 2 classification, there were credible effects of having a higher posterior probability of applying the norm of retraction to factual statements than for-all-we-know statements and might-statements, with 95% HDI of the contrast effects excluding zero. The only exception was for participants classified as following Objectivism who had received training in logic, where there were no credible effects of a difference between factual statements and might-statements (factual – might, $\tilde{\Delta} = .09$, 95% HDI [-.01, .19]).

A second finding was that the norm of retraction was applied more strongly to might-statements for participants classified as following Objectivism than for participants classified as following either Contextualism or Relativism. This was found for both participants who had not received prior training in logic (Contextualism – Objectivism, $\tilde{\Delta} = -.42$, 95% HDI [-.54, -.29], Objectivism – Relativism, $\tilde{\Delta} = .22$, 95% HDI [.10, .34]) as well as for participants who had received prior training in logic (Contextualism – Objectivism, $\tilde{\Delta} = -.19$, 95% HDI [-.36, -.02], Objectivism – Relativism, $\tilde{\Delta} = .24$, 95% HDI [.08, .40]).

Objectivists and relativists were predicted to apply the norm of retraction to might-statements, in contrast to contextualists (H_6). This prediction was followed by objectivists both with and without prior training in logic, but more strongly, if they had prior training in logic. In contrast, contextualists only followed the prediction of not applying the norm of retraction to might-statements if they had received no prior training in logic. Moreover, relativists did not follow the prediction concerning the norm of retraction as applied to might-statements. Accordingly, it was only for participants classified as adhering to Objectivism that prior training in logic facilitated following the theory's predictions. Hence, both (H_6) and (H_7) only received partial support and primarily from the participants classified as objectivists. On the other hand, (H_7) was challenged by the finding that participants classified as contextualists displayed more consistency with their assigned profile, if they had received no prior training

in logic. A further finding was that credible effects of applying the norm of retraction more strongly to might-statements than for-all-we-know statements for Objectivism and Relativism were not found. For Objectivism and Relativism such differences would, however, had been expected, given that only Contextualism treats the two types of statements as having the same meaning. Accordingly, (H₅) could not be supported by our results.

Summary of Findings from Session 1 and 2

In sum, in Experiment 1 it was found that there were individual differences in interpretations of epistemic modals in line with the re-analysis of published findings in the Introduction (H₁). In Session 1, it was found that participants both differed in their truth evaluations according to Relativism, Contextualism, and Objectivism, and that these truth evaluations were matched by incompatibility judgments underlying disagreement with epistemic modal statements. It was thereby shown that participants could be classified as following different latent classes of diverging interpretations of epistemic modals (H₄).

In each case, it was found that participants' truth evaluations could be distinguished from their evaluations of justification, which rules out a central alternative hypothesis of these findings (H₂). Finally, it was found that, of the three interpretations of epistemic modals, the one represented by Relativism was by far the most widespread. This in turn showed that the kind of shifts in truth value judgments of might-statements across different contexts of evaluation predicted by Relativism was frequently occurring (H₃).

Nevertheless, the findings in Experiment 1 also presented two puzzles. Contrary to (H₅), it was found that all latent classes of participants interpreted epistemic modals as similar in meaning to "for-all-we-know" statements in both session 1 and 2. Yet, such an equivalence is only sanctioned by Contextualism (see MacFarlane, 2014, pp. 265).

In addition, it was found in session 2 that the norm of retraction was applied to might-statements most strongly by participants, who had received training in logic and were classified as objectivists. Thus, only for these participants could a facilitation effect from prior

instruction in logic be supported (H₇). In contrast, the large subgroup of participants, who were classified as following Relativism showed no strong preferences towards applying the norm of retraction to epistemic modals, counter (H₆). This in turn poses a central explanatory challenge to the account of Relativism in MacFarlane (2014), which treats the norm of retraction as the main diagnostic criterion for separating Relativism from different variants of Contextualism.

Experiment 2

The goal of Experiment 2 was to reexamine these two puzzles from Experiment 1. First, to better implement the baseline, the time index of for-all-we-know statements was more clearly marked via the formulation “As of Monday, for all we know...”.⁹ By associating the updated information state with a specific day of the week, this manipulation permitted a simpler integration of the two time points in the for-all-we-know sentences themselves. With one group of participants, we tested the effects of these procedural changes, which leads to our first hypothesis for Experiment 2.

(H₈) Procedural changes to the baseline of for-all-we-know sentences should make participants less inclined to apply the norm of retraction to these statements.

Second, with a second group of participants, the diagnostic criteria from Bicchieri and Chavez (2013) and Bicchieri (2017) were adopted for testing the empirical reality of the norm of retraction. Bicchieri (2017) investigates ways to measure norms empirically and argues that social norms in general have multiple components: 1) they involve a conditional preference to comply by the norm, which 2) is guided by empirical and normative beliefs about whether other people follow the norm and whether the norm is sanctioned. Each of these components can be measured empirically. The empirical beliefs of participants concern whether the majority of people actually follow the norm in question and the associated risk of being

⁹ We are here grateful to Max Kölbel for this suggestion.

sanctioned by violations. The normative beliefs concern whether participants believe that the majority of people *should* follow the norm and whether norm violations *should* be sanctioned.

In Bicchieri and Chavez (2013), these diagnostic criteria were applied to measure participants' motivation for behaving in a self-serving fashion in a bargaining game. In Experiment 2, we applied some of these diagnostic criteria to examine whether a further group of participants believed that most participants thought that the norm of retraction should be applied to epistemic modal statements and what the probability of being sanctioned for its violation would be. This leads to our final research hypothesis.

(H₉) Bicchieri's diagnostic criteria for measuring norms match participants' reactions to direct violations of the norm of retraction.

Like Experiment 1, a baseline with factual statements was used, where the norm of retraction uncontroversially holds. Similarly, we also included prior logic training as a covariate to test for possible facilitation effects of prior logic training.

Method

Participants

The same sampling procedure and exclusion criteria were followed as Session 1 of Experiment 1. Unlike Experiment 1, Experiment 2 only had one session. A total of 292 people completed the experiment. After applying the *a priori* exclusion criteria, the final sample consisted of 207 people. Mean age was 41.19 years, ranging from 21 to 74. 50.73% of the participants self-identified as male, 48.79% self-identified as female, and one participant chose not to identify with either gender. 76.33% indicated that the highest level of education that they had completed was an undergraduate degree or higher. 30.92% indicated that they had received prior training in logic.

Design

The experiment had a mixed design with two factors. Sentence was a within-subject factor (with three levels: factual vs. might vs. know). The dependent variable (DV) factor was a mixed factor with levels distributed across two groups of participants (Group 1: DV_{agree} , Group 2: $DV_{sanction}$, $DV_{majority}$). To allow for six trial replications for each level of the Sentence factor, each participant went through 18 within-subject conditions in total.

Materials and Procedure

The same materials and procedures of Session 2 of Experiment 1 were applied to Group 1 in Experiment 2. An exception was that the continuation of the scenarios across t_1 and t_2 was split into two separate days (Monday vs. Wednesday) and that the for-all-we-know statements explicitly highlighted their time stamp via the formulation “As of Monday, FOR ALL WE KNOW, ...”.

The only difference between Group 1 and Group 2 was that instead of being asked whether they agreed/disagreed that the speaker should have taken the statement back (DV_{agree}), two new DVs based on Bicchieri (2017) were used.

The first ($DV_{majority}$) asked participants whether they agreed/disagreed with the statement that the majority of participants in the study think that the speaker should have taken back the claim by saying “Oh, then I guess I was wrong”. The second ($DV_{sanction}$) asked participants to rate on a scale from 0 to 100% the risk of being criticized for not taking back the statement (which was highlighted in blue to the participants).

Accordingly, the first dependent variable (DV_{agree}) was the same as in Session 2 of Experiment 1, but its formulation had changed. With one group of participants, we investigated these changes to the formulation of DV_{agree} . With a second group of participants, we presented the two new dependent variables from Bicchieri (2017). We decided to present these two new dependent variables in a separate group, to prevent carry-over effects of the dependent variable in the first group to these two new dependent variables.

Results and Discussion

Given the design, there were replicates for each participant and scenario. Accordingly, mixed-effects models, with random effects for intercepts of participants and of items were applied. This analysis was conducted using the statistical programming language R (R Core Team, 2015) and the R-package `brms` or mixed-effects models in Bayesian statistics (Bürkner, 2017). Since the accept and majority dependent variables were dichotomous, a Bernoulli likelihood function with a probit link function was used to model these variables (M_1). A Gaussian likelihood function was used for the continuous %-criticism dependent variable (M_2). Since the same group of participants had responded to both the majority and agree questions, the Sentence (factual vs. might vs. for all-we-know) and DV factors (agree vs. majority) were allowed to interact for this group. Both models included random intercepts and slopes for participants and items.

Following Experiment 1, we also fitted a pair of models that included interactions based on whether participants had previously received instruction in logic. However, unlike in Experiment 1, we did not in general find credible effects for interactions with prior training in logic and the model fitting criteria did not indicate an improvement in the fit by including the interactions.¹⁰ We therefore decided in favor of the models without the interaction effects with prior training in logic and display their posterior predictions in Figure 7.

¹⁰ M_1 without Sentence * DV * logic interaction (LOOIC = 3272.0); M_1 with Sentence * DV * logic interaction (LOOIC = 3273.2). M_2 without Sentence*logic interaction (LOOIC = -653.0); M_2 with Sentence*logic interaction (LOOIC = -653.4).

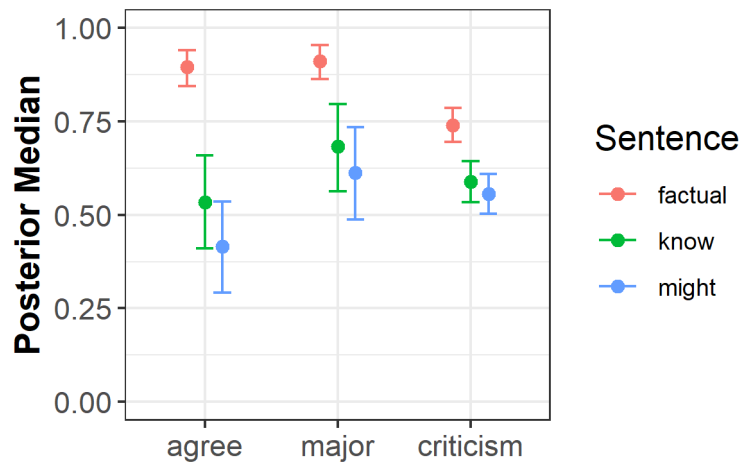


Figure 7. *Norm of Retraction*. Across various sentence types, Figure 7 displays the posterior predicted median probabilities of 1) agreeing that breaching the norm of retraction is a norm violation (“agree”), 2) judging that the majority of participants would think that breaching the norm of retraction is a norm violation (“major”), and 3) the risk of being criticized for failing to comply with the norm of retraction (“criticism”). The error-bars represent 95% CI intervals.

For all three dependent variables, there was a tendency for the ratings to follow the following rank-ordering: factual > know/might with 95% HDI for the contrast effects excluding 0. In addition, it was found that the data did not support a difference between might-statements and for-all-we-know statements for judgments about the majority of other participants (know – might, $\tilde{\Delta} = .07$, 95% HDI [-.02, .16]) and the risk of being criticized (know – might, $\tilde{\Delta} = .03$, 95% HDI [.00, .06]). For judgments about agreeing that breaching the norm of retraction is a norm violation slight differences emerged (know – might, $\tilde{\Delta} = .12$, 95% HDI [.03, .21]).

The results displayed in Figure 7 are in line with the findings of Experiment 1, but the explicit inclusion of the time index “As of Monday” in the for-all-we-know statements lead to a better implementation of the baseline (H_8). For Objectivism and Relativism, it poses a problem that the central tendency from Experiment 1 could be replicated that participants lack strong preferences for applying the norm of retraction to might-statements.

What Experiment 2 adds is that this lack of inclination was matched by their judgments about whether violations are sanctioned and whether the majority of participants think that the norm of retraction should be applied to might-statements (H_9). Under the assumption that there is a norm of retraction for might-statements, we would have expected

that diagnostic criteria for the presence of norms like Bicchieri's (2017) would have indicated that participants applied the norm of retraction to a similar extent to might-statements as to the baseline of factual statements. What was found instead was that participants did not apply the norm of retraction to a stronger degree to might-statements than to for-all-we-know statements, which is a case where the norm is commonly thought not to apply.

In Experiment 2, strong evidence in favor of interactions with prior logical training were not found. In Experiment 1 such interactions were found. But we also saw that these interactions depended on which interpretation of might-statements that participants had been assigned to, which is a factor that the simpler experimental design of Experiment 2 did not investigate. Thus, abstracting away from latent profile of interpretation of might expressions, Experiment 2 suggests that no general facilitation of performance with the norm of retraction based on prior training of logic was found, contrary to (H₇).

General Discussion

Truth-conditional semantics developed into a flourishing theoretical framework by accounting for the compositionality of sentences describing the outer world. Eventually, it was enriched to handle context sensitivity (Kaplan, 1989) and counterfactual alternatives to the course of history (Lewis, 1973). One aspect that escaped it was, however, subjective language as involved in taste judgments and uncertainty (Lasersohn, 2017). In a recent discussion, several theoreticians have proposed to extend the toolkit of formal semantics with relative truth values to fill this gap (Egan, Hawthorne, & Weatherson; Kölbel, 2009, 2015a, 2015b; Egan, 2007; MacFarlane, 2011, 2014; Lasersohn, 2017). In MacFarlane (2011, 2014), this takes the form of making the truth values of a range of expressions (e.g., conditionals, subjective taste predicates, epistemic modals, and knowledge ascriptions) relative to a context of assessment.

In contrast, one of the main theoretical alternatives, Contextualism, makes the truth value of sentences containing epistemic modals dependent on the information state at the

context of utterance (Hacking, 1967; DeRose, 1991; Kratzer, 1977, 2012; von Fintel & Gillies, 2008, 2011). On this view, epistemic modals are used to make statements *about* a particular body of evidence available at the context of utterance (c_1). For Relativism, on the other hand, the content of the statements expressed is not about any body of evidence in particular but has a truth value that varies depending on the evidence available at the context in which it is assessed, c_2 (Khoo & Phillips, 2018).

A key diagnostic case for arbitrating in this dispute is eavesdropping cases (MacFarlane, 2011). By placing the speaker and the interlocutor at contexts with diverging information states, these cases dissociate the influence of the context of utterance from that of the context of assessments (Katz & Salerno, 2017). As such, eavesdropping cases directly address the need for semantic values that are a function of contexts of assessments. What's more, eavesdropper cases characterize a situation in which the interlocutor knows more than the speaker, and yet chooses to evaluate the speaker's claim for its truth/falsity in the light of information that is only accessible to the interlocutor.

If it could be empirically substantiated that competent English speakers evaluate statements containing epistemic modals in this way, then it would constitute a direct challenge to the core claim of the contextualist that only the information available at the context of utterance matters for the truth of epistemic modals. As Katz and Salerno (2017, p. 142) note:

Movement from the default contextualist position to a more exotic relativist framework then requires forceful motivation. In that spirit the relativist emphasizes empirical data that allegedly only she can accommodate.

Consequently, it was investigated in several recent papers whether empirical data supports such a shift from Contextualism to Relativism. For instance, in Knobe and Yalcin (2014), it was investigated whether Relativism adequately characterizes data on truth values and retraction. In Katz and Salerno (2017) and Khoo and Phillips (2018), it was investigated whether Relativism adequately characterizes data on the compatibility/incompatibility of two

contrary truth value assignments. In each case, it was found that the evidence in favor of Relativism was lacking. But the empirical picture was complicated by the fact that the results did not support standard Contextualism either.

We have already seen how some of this evidence relies on reporting statistical analyses at the group level. When re-examining their results, it was found that they contained considerable variation at the individual level, as we have seen. Consequently, in the present paper we set out to investigate individual variation in the interpretation of epistemic modals by conducting a study with two test sessions. In the first, participants were classified according to three semantic interpretations of epistemic modals at the individual level based on both their truth value assignments. In the second, their adherence to the norm of retraction was investigated. For the remainder, we present some of the open questions raised by our results. We will focus on argumentation with epistemic modals. Finally, we consider how our results relate to other theories in psychology that deal with epistemic modals.

Shifting Truth Values

Historically, the notion of relative truth has had a bad reputation in that it became mingled up with general debates about postmodernism, the science wars, and social constructivism (Boghossian, 2007). Indeed, the very notion has often been suspected of being circular, although the details of the argument are not that simple (Nozick, 2001). It took technical refinements to show that a respectable formal notion of relative truth could be explicated via relativity to contexts of assessment (MacFarlane, 2014). This enabled use of the notion as a theoretical construct to account for a fragment of natural language. But the empirical adequacy of this proposal still needs to be established, since the norms in question are supposed to account for our linguistic competence.

The radicality of introducing relative truth values for this purpose consists in permitting both that the truth values of the same content can diverge between multiple

interlocutors and can shift across time for the same interlocutor. Since the content expressed stays invariant, this notion is importantly different from the more common notion that the same sentence can express different contents in the mouth of different speakers (MacFarlane, 2014). One salient example is sentences containing indexicals and other context-sensitive terms (e.g., “I am hungry”). For the contextualist, sentences containing epistemic modals (e.g., *might*, *must*) express a similar kind of context sensitivity. In contrast, for the relativist, such sentences can express the same content across different contexts and yet vary in truth values for different speakers as well as shift truth values for the same speaker across time. As emphasized by Wright (2008), it is crucial that direct evidence be had for shifting truth values across different context of assessment for the empirical case for Relativism. Through our experimental investigations, we have been able to supply some of the first empirical evidence for shifting truth values. We saw this in Figure 4, where Relativists had a posterior median above 80% of finding the *might*-statements true before *not-p* became known and a posterior median close to zero of finding them true after *not-p* became known.

MacFarlane (2014, p. 107) is skeptical of the prospect of seeking direct evidence for Relativism in this way. The reason is that the semantic theory of Relativism employs a technical notion of truth that applies to utterances at contexts, whereas he takes it that ordinary speakers use a monadic truth predicate that only applies to the propositional content expressed by sentences. In contrast, Kölbel (2015b, pp. 7) puts forward a principle that connects relativistic semantics with measurable data, which presupposes that competent language users *can* judge the truth of potential utterances of sentences. In Experiment 1 (phase 1), we attempted to collect such data by referring to a token statement made at t_1 and asking participants whether that statement *was* true when it was made and whether the statement made at t_1 *is* true after the continuation had been learned at t_2 (see Table 4). Using this set-up, we were able to find qualitatively distinct patterns in participants' truth assignments matching the predictions of the competing theories.

Both Khoo and Phillips (2018) and Cantwell (forthcoming) consider the possibility that mixed results in previous studies may have been facilitated by a difficulty in separating truth and justification by the participants. We therefore measured whether participants thought that the speaker was justified in making the statement as a control in Experiment 1 to test this alternative hypothesis. The results indicate that irrespectively of which semantic interpretation participants adopt of epistemic modals, they are capable of distinguishing between truth and justification for the various sentences examined.

Against prior expectations, it was, however, found that for each of these three interpretations, the evaluation of for-all-we-know statements and epistemic modals were aligned. *A priori* such an alignment would only have been expected for participants classified as following Contextualism (MacFarlane, 2011, 2014). Yet, it was found that participants differed as sharply in their interpretations of for-all-we-know statements as they did for epistemic modals. One hypothesis is that a clearer demarcation of the time indices of the points of evaluation would help participants to draw a distinction between the two types of statements. Partial support for this hypothesis could be obtained in Experiment 2, where t_1 and t_2 were separated into distinct days and the for-all-we-know statements included the explicit qualification “As of Monday, for all we know...”. The support for this was only partial, because credible differences were found only for judgments about agreeing that breaching the norm of retraction is a norm violation but not for the further diagnostic criteria based on Bicchieri (2017) employed in Experiment 2.

In Experiment 1, it was found that participants classified as following Relativism and Objectivism extended their tendency to evaluate past tokens of “for all we know, p ” as asserted by a speaker at t_1 (14:10), by their own present state of knowledge at t_2 (14:12), just as they would with might-statements. Possibly, these participants were misled by that “might p ” sounds similar in meaning to “for all we know, p ” when presented interchangeably in a

within-subject design, although their truth evaluation of the former did not fit for the latter. Future studies will have to look into this possibility.¹¹

Argumentation with Epistemic Modals

A large part of the debate over Relativism concerns argumentation with subjective expressions such as epistemic modals. Views accordingly differ on whether a Trump supporter uttering “Trump might have won the 2020 election” and a Democrat denying this are in genuine disagreement, or whether they are merely talking past one another by expressing features of their own subjective state of uncertainty (see the essays in Egan & Weatherson, 2011). A central motivation for Relativism is to allow that two speakers may be in a state of genuine disagreement about the same content but where each is correct according to their own perspective (Kölbel, 2009).

At the same time, Relativism aims at securing a basis for argumentation via its norm of retraction (MacFarlane, 2011, 2014): although the truth values of statements containing epistemic modals change with the information state of the context of assessment, statements that come out as false at any given time need to be retracted. Consequently, if the Trump-supporter above changes her information state by taking the outcome of the 2020 election at face value, then her earlier might-statement would have to be retracted. Yet, if the information states of neither the Democrat nor the Trump-supporter changes, they may continue to be in a state of “faultless” disagreement as far as the might-statement goes (Kölbel, 2009).

When examining the norm of retraction in Experiments 1 and 2, high-stakes scenarios were used where a speaker makes a public statement in, e.g., media outlets, the court, or in

¹¹ A further possibility includes using ‘for all I know’ instead ‘for all we know’ as a baseline. We originally decided against this option, since the position of Solipsistic Contextualism, which narrows the context down to just consist of the speaker, is rarely advocated. Rather, the more wide-spread view is that the context includes the joint information state of a group of interlocutors (MacFarlane, 2014). Yet, results in Kneer (forthcoming) indicate that this other baseline would have been easier to implement.

scientific disputes. Still, it was found that participants in general did not have strong inclinations to apply the norm of retraction to epistemic modals, contrary to the predictions of Relativism and Objectivism.

After having examined patterns of individual variation, we can thus agree with Knobe and Yalcin (2014) that strong support for the norm of retraction as applying to epistemic modals is, in general, not found. But it was, however, found that the shifting truth evaluations of Relativism characterized the largest latent class of participants in Experiment 1, and that these participants consistently stood by these judgments when challenged by contrary views in the Scorekeeping task. Moreover, it was found that these truth evaluations were matched within this sub-group by corresponding incompatibility judgments. So, whereas Katz and Salerno (2017) and Khoo and Phillips (2018) find at the group-level that participants do not follow the predictions of Relativism of treating two contrary epistemic modal statements with different evidence accessible as incompatible, support for this prediction was found in Experiment 1 for the largest subgroup of participants.

Based on our results, it is likely that ordinary people on average do not apply the norm of retraction to epistemic modals, because they regard them as a type of hedged assertions, which are less costly to make. Williamson (2020, p. 11) gives expression to this intuition:

When the detective rightly asserts ‘Smith must be guilty’, she does not regard her earlier assertion of ‘Smith may be innocent’ as in any way wrong. It was not a *mistake* made on the basis of incomplete but misleading evidence. That was the point of saying ‘Smith may be innocent’ rather than ‘Smith is innocent’.

In support, it was found in Experiment 2 that participants tend not only to lack strong preferences to apply the norm of retraction to epistemic modals themselves, but that they also accurately judge that the majority lack similar preferences, and that the chances are low for being sanctioned, if one refused to retract a previous might-statement. So, while the norm of retraction provides an interesting normative justification in MacFarlane (2011, 2014) for

Relativism as applied to epistemic modals – in terms of the foundation for argumentation it provides – it is not a norm that participants in general have strong normative views about for might-statements. It is only for factual statements that prescriptions of retraction appear to be an important norm to most of the participants we investigated.

However, our study of individual differences permits us to qualify this negative assessment in one important aspect. The results indicate that for participants who have had training in logic and committed to Objectivism in the Scorekeeping task, the norm of retraction was applied to the same degree to might-statements as factual assertions. So at least for this minority, the norm of retraction finds application to epistemic modals.

Our findings leave us with a general explanatory challenge: how is argumentation with epistemic modals possible, given that it is not based on the norm of retraction as MacFarlane (2014) has argued? Since epistemic modals are used to state alternative hypotheses in science, it would be strange if no norms of argumentation applied to them.

Addressing this first challenge is further complicated through a second challenge raised by another topic we encountered. Based on the results of Experiment 1, as well as our re-analysis of published findings, we have found grounds for questioning the following widely shared assumption:

(U) There is a uniform interpretation of expressions like epistemic modals. Only one of conflicting semantic theories can be descriptively adequate. If semantic theories like Contextualism, Relativism, and Objectivism are incompatible, then at most one of them can be descriptively correct.

Rejecting (U) implies that to the extent that argumentation over epistemic modals occurs, it is complicated by the circumstance that different speakers may apply different interpretations for their semantic evaluation. This lack of a semantic foundation for a shared standard does not necessarily render argumentation over epistemic modals impossible, just as people can engage in argumentation over moral judgments despite wide disagreement on underlying

moral standards. But it likely renders argumentation over epistemic modals more fragile and complicates its theoretical explication.

Nevertheless, it is possible to identify components of argumentation with epistemic modals on which contextualists, relativists, and objectivists can all agree. In Appendix 1, Table A1 we present an overview over such cases, which shows that there are cases which regulate which assertions can be made, and which challenges are appropriate, on which interlocutors can agree, even if they differ on the underlying interpretation of epistemic modals. For instance, if p is not excluded by the shared information state, relativists, contextualists, and objectivists can all agree that “might- p ” is assertable and that the speaker is warranted in rejecting challenges to her assertion of might- p . Moreover, the interlocutor can use factual information to challenge the speaker to update the shared information state so that p is excluded and might- p is not warrantably assertable. *Vice versa*, if the shared information state had already excluded p at the outset, then might- p would not be assertable. In which case, interlocutors can warrantably challenge assertions of might- p , or alternatively, use factual evidence to challenge the speaker to update the shared information state so that might- p becomes assertable.

Thus, despite the uncertainty of whether one is communicating with a person who evaluates might-statements according to Contextualism, Relativism, or Objectivism, it is possible to follow these rules for arguing over might-statements and to coordinate which challenges need to be addressed by the speaker. By following these rules, the Democrat and the Trump-supporter can argue over whether it is correct to assert “Trump might have won the 2020 election”. But as this example illustrates, often the root of the dispute is not so much the might-statement itself but a dispute about whether an update to the underlying information state with factual information should be performed. In their case, the root of the dispute concerns factual information about the 2020-election.

What this example illustrates is that there is an important asymmetry between arguing over epistemic modals and arguing over other types of subjective expressions, like taste judgments. The discourse over epistemic modals parasitizes on factual discourse, whereas taste judgments only have subjective standards of taste to rely on. For this reason, argumentation with epistemic modals have the additional resources of being able to dissolve into a dispute over which updates with factual information are mandated. For taste judgments, this option is absent and for this reason a norm of retraction of the kind that MacFarlane (2014) introduces *may* prove more important for argumentation over taste judgments than for argumentation with epistemic modals.¹²

As part of their argument in favor of the norm of retraction, MacFarlane (2014) and Cantwell (forthcoming) interpret the norm as not implying an admission that the speaker was at fault for having made the assertion at t_1 . Rather, the retraction is taken as an admission that the statement made at t_1 is semantically false at t_2 after *not- p* has become known. The first type of assessment concerns how well the speaker performed in using the might expression; the second concerns the correctness of the *content* uttered (Kölbel, 2015a, fn. 5). This distinction is subtle, and it therefore remains to be seen whether future studies can find stronger support for the norm of retraction, given other formulations of retractions.

Since epistemic modals have played a role in a recent controversy concerning probability and mental model theory concerning epistemic possibilities as the foundation of psychology of reasoning (Hinterecker et al. 2016, Johnson-Laird & Ragni, 2019, Oaksford, Over et al., 2019), we finally consider relations between these theories and our results.

Comparisons to Theories in Psychology

On popular Bayesian approaches to reasoning, generalized quantifiers (e.g., ‘most’) have been analyzed probabilistically (Oaksford & Chater, 2007). Complementary, Lassiter

¹² Results in Kneer (2021) cast doubt on this possibility, however.

(2017) has analyzed ‘possible’ as a gradable, scalar concept, which can be modeled probabilistically, instead of being an all-or-nothing notion captured by existential quantification. On this account, epistemic modals and probabilities share a common ratio scale. As a possible ordering of epistemic modals, Lassiter (p. 152) considers the following order of probabilistic thresholds ($p(\varphi) > \theta_i$):

$$\theta_{possible} < \theta_{might} < \theta_{likely} < \theta_{must} < \theta_{certain}$$

Normally, ‘possible’ and ‘might’ are treated as equivalent (e.g., with $\theta_{possible} = \theta_{might} = 0$), but Lassiter (2017) presents experimental evidence indicating that the former is weaker than the latter (i.e., $\theta_{might} > 0$). This observation is important because the main theory of modal reasoning in psychology (i.e., mental model theory) has explicated a semantics for ‘possible’ (Johnson-Laird & Ragni, 2019), but not directly for ‘might’.

In this paper, we have focused on a single epistemic modal without considering issues of gradeability and varying strength of probabilistic information. A natural follow-up would therefore be to investigate whether qualitatively distinct patterns of individual differences persist once further epistemic modals are tested, and the strength of the probabilistic evidence is varied. On the other hand, a scalar semantics for epistemic modals would also have to account for our data concerning truth value judgments. On a scalar semantics, binary truth values can be recovered via the thresholds; values above the threshold corresponding to ‘true’ and those below corresponding to ‘false’ (Lassiter, 2017, Ch. 1). The separation between truth and justification in our data could then be accounted for by holding that participants’ judgments of justification are based on probabilities and probabilistic thresholds introduce binary distinctions between true and false. But now consider the puzzle raised by our results.

For some participants, there was a high probability that *might-p* was justified at t_1 , *might-p* is evaluated as having met the threshold at t_1 , and continues to be evaluated as meeting the threshold at t_2 , although *not-p* is known as a fact (contextualists). For other participants, there was a high probability that *might-p* was justified at t_1 , *might-p* is evaluated

as having met the threshold at t_1 , and *might-p* is evaluated as not meeting the threshold at t_2 (relativists). For yet other participants, there was a high probability that *might-p* was justified at t_1 , *might-p* is evaluated as not having met the threshold at t_1 , and continues to be evaluated as not meeting the threshold at t_2 (objectivists). Since *might-p* retains a high probability of justification for all participants, the participants would have to be modelled as differing in whether the threshold for binary truth judgements is fixed by the information state at t_1 (contextualists), it shifts between t_1 and t_2 to track the context of assessment (relativists), or whether it is fixed by the known facts at t_2 (objectivists). While it is conceivable that the scalar semantics of Lassiter (2017) could be extended with a theory of individual variation in thresholds and their context-dependence on information states, it is also clear that the theory would need to be extended in some way to be able to account for our results.

As Lassiter (2017, pp. 5-6, 157) notes, no one has yet a complete theory about the context-sensitive interpretations of scalar expressions in general; including a complete theory of the contextual factors that influence the probabilistic thresholds of epistemic modals. But the challenge from our results is different. It suggests that for different participants, different information states influence how the thresholds are set in the same context, if a probabilistic threshold-account of our data on truth value judgments is to be viable.

Turning to mental model theory, we find a very different outlook on epistemic modals. Like Kratzer (2012), mental model theory focuses on ordinal comparisons between epistemic modals. The theory posits that reasoning depends on models of possibilities. Accordingly, mental model theory presupposes that participant parse natural language into representations of the possibilities that the sentences assert, as illustrated in Table 13. For compound sentences (e.g., ‘if p then q ’), the semantic meaning is explicated via a conjunction of possibilities, and so if participants are to fully process it (and not take heuristic shortcuts like only processing the first model), they should arrive at all the conjuncts listed.

Table 13. Parsing of Natural Language in MMT

Statements	Possibilities added to the Mental Model	Status
p	$\diamond p$	Fact
p and q	$\diamond(p, q)$	Fact
p or q	$\diamond(p, q) \wedge \diamond(p, \neg q) \wedge \diamond(\neg p, q)$	Epistemic possibilities
If p then q	$\diamond(p, q) \wedge \diamond(\neg p, q) \wedge \diamond(\neg p, \neg q)$	Epistemic possibilities
It is possible that p	$\diamond p \wedge \diamond \neg p$	Epistemic possibilities
It is not possible that p	$\diamond \neg p$	Fact

Note. The table states the fully explicit models in logical notation. In addition, mental model theory also has a performance theory for heuristic processing, where participants only consider the first possibility. ‘ $\diamond p$ ’ = it is possible that p . ‘ $\diamond(p, q)$ ’ = it is possible that p and q . ‘ $\Box p$ ’ = it is necessary that p . ‘ \wedge ’ = logical conjunction. For comparison, see e.g., Johnson-Laird and Ragni (2019, Table 2). When only one possibility is added to the mental model of the premises by a statement, this possibility acquires the status of being a fact (Khemlani et al., 2018).

These conjunctions are thought to be exhaustive and so every other combination is treated as impossible. Factual statements, like ‘ p ’ and ‘ p and q ’, only assert one possibility (Khemlani et al., 2018). Recently, mental model theory has been extended to apply to sentences containing epistemic modals (Johnson-Laird & Ragni, 2019; Ragni & Johnson-Laird, 2020). Accordingly, ‘it is possible that p ’ presupposes the following conjunction of possibilities: ‘ p is possible’ and ‘not- p is possible’.

Could mental model theory account for our results? In the absence of a direct account of might-statements, we use what mental model theory says about ‘possible’ as a proxy, since traditionally the two have been assumed to be similar in meaning and Ragni and Johnson-Laird (2020) treat ‘may’, ‘might’, and ‘could’ as all indicating possibility in their experiments. As mental model theory is silent on issues pertaining to eavesdropper cases and the distinction between context of use and context of assessment, we will have to extrapolate to apply it to the Eavesdropper Task.

For this, the reader is encouraged to think back at the temporal structure in Table 4. When presented with the fact that not- p in this task, it necessarily follows that ‘it is not possible that p ’ on mental model theory. Indeed, the two are equivalent on the revised theory,

because they each only have one explicit model ($\diamond\neg p$).¹³ Since, moreover, ‘it is possible that p ’ presupposes both ‘ p is possible’ and ‘not- p is possible’ on mental model theory, a presupposition of ‘it is possible that p ’ is violated at t_2 when not- p has become known. Whether the violation of this presupposition merely renders the statement unassertable at t_2 (as all theories agree), a truth-value gap, or false (as Relativism and Objectivism would hold), is not clear.

To accord with the modal tendency of our results corresponding to the relativistic response pattern, mental model theory would have to hold that ‘it is possible that p ’ as asserted in t_1 is made false by the disclosure of not- p at t_2 . Accordingly, to capture the relativistic response, mental model theory would have to be extended by the auxiliary assumption that the violation of said presupposition results in a ‘false’ evaluation at t_2 . Yet, mental model theory would still be challenged by the finding that there are other participants (classified as contextualists), who continue to treat the might-statement as true, despite the violated presupposition. Moreover, the difference between relativists and objectivists in whether the might-statement was considered true or false at t_1 is left unaccounted for. Comments in Ragni and Johnson-Laird (2020) on the voidness of a claim with a false presupposition could indicate that their theory predicts a truth-value gap, in which case they would have to predict that the modal tendency of giving a ‘false’ evaluation at t_2 would be reversed on a ternary truth value format.

¹³ Due to the duality of $\diamond p$ and $\Box\neg p$ in modal logic, ‘it is not possible that p ’ ($\neg\diamond p$) would normally be equivalent to ($\Box\neg p$). The revised mental model theory gives up this duality for epistemic modality but retains it for deontic modality (Ragni & Johnson-Laird, 2020) and appears to have difficulties distinguishing between the explicit models of factual statements (p) and of the much stronger statements of necessity ($\Box p$). Perhaps for this reason, Johnson-Laird and Ragni (2019) offer an idiosyncratic reinterpretation according to which a single statement of a necessity states a necessary condition for another state of affairs (p. 11). But this account is liable to run into troubles when it comes to representing the strongest epistemic modals (e.g., ‘ p is certain’) and it misrepresents the content of factual statements. The problem arises due to mental model theory’s modalization of descriptive sentences (see also Over, 2022).

To further apply mental model theory to account for all our results runs into the difficulty that mental model theory has not explicitly been developed to address the same questions as the semantic theories investigated in our experiments (i.e., eavesdropper cases, relativity of truth values to context of use and context of assessment, the distinction between truth and justification for might-statements, retractions, and argumentation with might-statements). As such, our results present an opportunity for both probabilistic approaches and mental model theory in psychology to expand their theories to cover new domains in competition with the cluster of theories we investigated. Since neither theory predicted our results, the challenge is whether the theories can be extended by suitable auxiliary hypotheses that can lead to novel findings of their own.

Conclusion

In this paper, we have investigated individual variation in the adherence to Relativism, Contextualism, and Objectivism about epistemic modals. Our experimental investigations were motivated by a re-analysis of existing empirical data (Knobe & Yalcin, 2014; Khoo & Phillips, 2018), which we showed contained a substantial degree of individual variation which is not captured by statistics for central tendencies at the aggregated group level.

As an alternative strategy, we followed the individual classification approach of Skovgaard-Olsen et al. (2019), which had previously only been applied to conditionals. On this approach, latent profiles of participants' case judgments and reflective attitudes are established. It was then probed whether participants are internally consistent across multiple sessions when comparing their latent profile to their adherence to the consequential norms. In the first session of Experiment 1, it was found that assignments to latent classes of interpretations of might-statements could be used to predict differences in participants' truth value assignments in so-called eavesdropping cases as well as their evaluations of disagreements. Through these classifications, we were able to obtain some of the first

empirical evidence for shifts in truth values across different time points, in agreement with Relativism. This in turn provides an empirical validation for the theoretical notion of content with relative truth values used as a semantic device for modelling fragments of natural language in MacFarlane (2014).

In a second session, it was then tested whether these three semantic interpretations were consistent vis-à-vis their performance with the norm of retraction informing debates between Contextualism and Relativism. With the exception of participants who had received training in logic and committed to Objectivism, it was found that none of the groups showed a strong adherence to the norm of retraction for epistemic modals. Yet, this norm provides one of the most central motivations for Relativism in MacFarlane (2014). The results of Experiment 2 corroborated this conclusion of lack of general adherence to the norm of retraction for epistemic modals by applying Bicchieri's (2017) diagnostic criteria.

References

- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Beddor, B. and A. Egan (2018). Might do better: Flexible relativism and the QUD. *Semantics and Pragmatics*, 11(7).
- Bicchieri, C. (2017). *Norms in the Wild. How to Diagnose, Measure, and Change Social Norms*. Oxford: Oxford University Press.
- Bicchieri, C. and Chavez, A. K. (2013). Norm Manipulation, Norm Evasion: Experimental Evidence. *Economics and Philosophy*, 29, 175-98.
- Boghossian, P. A. (2007). *Fear of Knowledge: Against Relativism and Constructivism*. Oxford: Oxford University Press.
- Bürkner, P. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, 80(1), 1-28.

- Bürkner, P., and Vuorre, M. (2018, February 28). Ordinal Regression Models in Psychological Research: A Tutorial. Retrieved from <http://doi.org/10.17605/OSF.IO/X8SWP>
- Cantwell, J. (forthcoming). Objective epistemic modals. (*in review*)
- Cole, R. P., Barnet, R. C., and Miller, R. R. (1997). An Evaluation of Conditioned Inhibition as Defined by Rescorla's Two-Test Strategy. *Learning and Motivation*, 28, 323-341.
- DeRose, K. (1991). Epistemic Possibilities. *Philosophical Review*, 100(4), 581–605.
- Egan, A. (2007). Epistemic Modals, Relativism, and Assertion. *Philosophical Studies*, 133(1), 1–22.
- Egan, A., J. Hawthorne, and Weatherson, B. (2005). Epistemic Modals in Context. In G. Preyer and P. Peter (Eds.), *Contextualism in Philosophy* (pp. 131–170). Oxford: Oxford University Press.
- Egan, A. and Weatherson, B. (Eds.) (2011). *Epistemic Modality*. Oxford: Oxford University Press.
- Elqayam, S. and Evans, J. St. B. T. (2011). Subtracting “ought” from “is”: descriptivism versus normativism in the study of human thinking. *Behavioral and Brain Sciences*, 34, 233-90.
- Evans, J. S. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin*, 128(6), 978-996.
- Farell, S. and Lewandowsky, S. (2018). *Computational Modeling of Cognition and Behavior*. New York: Cambridge University Press.
- Hacking, I. (1967). Possibility. *Philosophical Review*, 67, 143–168.
- Heim, I. and Kratzer, A. (1998). *Semantics in Generative Grammar*. Oxford: Blackwell Publishing.

- Hinterecker, T., Knauff, M., and Johnson-Laird, P. N. (2016). Modality, Probability, and Mental Models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(10), 1606-1620.
- Johnson-Laird, P. N., & Ragni, M. (2019). Possibilities as the foundation of reasoning. *Cognition*, 193, Article 103950.
- Kaplan, D. (1989). Demonstratives. In Almog, J., Perry, J., and Wettstein, H. K. (Eds.). *Themes from Kaplan* (pp. 481-563). Oxford: Oxford University Press.
- Katz, J. and Salerno, J. (2017). Epistemic Modal Disagreement. *Topoi*, 36, 141-153.
- Kellen, D. and Klauer, K. C. (2018). Elementary signal detection and threshold theory. In E. J. Wagenmakers (Ed.), *Stevens' Handbook of Experimental Psychology and Cognitive neuroscience (4th edition, vol. v)*. New York: Wiley.
- Khoo, J. and Phillips, J. (2018). New horizons for a theory of epistemic modals. *Australian Journal of Philosophy*, 97(2), 309-324.
- Klauer, K. C. (2010). Hierarchical Multinomial Processing Tree Models: A Latent-Trait Approach. *Psychometrika*, 75(1), 70-98.
- Klauer, K. C., Beller, S., and Hütter, M. (2010). Conditional Reasoning in Context: A Dual-Source Model of Probabilistic Inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(2), 298-323.
- Kneer, M. (2021). Predicates of personal taste: empirical data. *Synthese*, 199, 6455-6471.
- Kneer, M. (forthcoming). Epistemic Modal Claims – Data and ‘data’.
- Knobe, J. and Yalcin, S. (2014). Epistemic modals and context: Experimental data. *Semantics & Pragmatics*, 7(10), 1-21.
- Kratzer, A. (1977). What ‘must’ and ‘can’ must and can mean. *Linguistics and Philosophy*, 1(3), 337–356.
- Kratzer, A. (2012). *Modals and Conditionals: New and Revised Perspectives*. Oxford: Oxford University Press.

- Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.
- Kölbel, M. (2003). Faultless Disagreement. *Proceedings of the Aristotelian Society*, 104, 53-73.
- Kölbel, M. (2009). The Evidence for Relativism. *Synthese*, 166(2), 375-95.
- Kölbel, M. (2015a). Relativism 1: Representational Content. *Philosophy Compass* 10(1), 38-51.
- Kölbel, M. (2015b). Relativism 2: Semantic Content. *Philosophy Compass* 10(1), 52-67.
- Lasersohn, P. (2017). *Subjectivity and Perspective in Truth-Theoretic Semantics*. Oxford: Oxford University Press.
- Lassiter, D. (2017). *Graded Modality: Qualitative and Quantitative Perspectives*. Oxford: Oxford University Press.
- Lee, M. D., and Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge: Cambridge University Press.
- Lenth, R. (2020). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.5.1. <https://CRAN.R-project.org/package=emmeans>
- Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell Publishing.
- Li, Y., Lord-Bessen, J., Shiyko, M., and Loeb, R. (2018). Bayesian Latent Class Analysis Tutorial. *Multivariate Behav Res.*, 53(3), 430-451.
- Linzer, D. A., and Lewis, J. B. (2011). poLCA: An R Package for Polytomous Variable Latent Class Analysis. *Journal of Statistical Software*, 42(10), 1-29.
- MacFarlane, J. (2011). Epistemic Modals Are Assessment-Sensitive. In B. Weatherson and A. Egan (Eds.), *Epistemic Modality* (pp. 144–178). Oxford University Press.
- MacFarlane, J. (2014). *Assessment Sensitivity: Relative Truth and its Applications*. Oxford: Oxford University Press.

- Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *Journal of Open Source Software*, 4(40), 1541. [10.21105/joss.01541](https://doi.org/10.21105/joss.01541)
- Miller, R. R., Hallam, S. C., Hong, J. Y., & Dufore, D. S. (1991). Associative structure of differential inhibition: Implications for models of conditioned inhibition. *Journal of Experimental Psychology: Animal Behavior Processes*, 17, 141–150.
- Mair, P. (2018). *Modern Psychometrics with R*. Cham: Springer.
- Nozick, R. (2001). *Invariances: The Structure of the Objective World*. Cambridge, MA: Harvard University Press.
- Oaksford, M., and Chater, N. (2007). *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. Oxford: Oxford University Press.
- Oaksford, M., Over D. E., & Cruz, N. (2019). Paradigms, possibilities, and probabilities: Comment on Hinterecker et al. (2016). *Journal of Experimental Psychology: Learning Memory and Cognition*, 45, 288-297.
- Over, D. E. (2022). The new paradigm and massive modalization. *Thinking & Reasoning*. <https://doi.org/10.1080/13546783.2021.2017346>
- Plummer, M. (2019). *rjags: Bayesian graphical models using MCMC*. R package version 4-10. URL = <https://CRAN.R-project.org/package=rjags>.
- Portner, P. (2009). *Modality*. Oxford: Oxford University Press.
- R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Ragni, M. and Johnson-Laird, P. N. (2020). Reasoning about epistemic possibilities. *Acta Psychologica*, 208, 103081.
- Skovgaard-Olsen, N. (2019). The Dialogical Entailment Task. *Cognition*, 193.
- Skovgaard-Olsen, N., Kellen, D., Hahn, U., and Klauer, K. C. (2019). Norm Conflicts and Conditionals. *Psychological Review*, 126(5), 611-633.

- Teller, P. (1972). Epistemic possibility. *Philosophia*, 2(4), 303–320.
- Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*. 27(5), 1413-1432.
- von Fintel, K. and A. Gillies (2008). CIA Leaks. *Philosophical Review*, 117(1), 77–98.
- von Fintel, K. and A. Gillies (2011). ‘Might’ Made Right. In A. Egan and B. Weatherson (Eds.), *Epistemic Modality* (pp. 108-130). Oxford: Oxford University Press.
- Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., et al. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychon Bull Rev*, 25, 35-57.
- Weatherson, B. (2009). Conditionals and Indexical Relativism. *Synthese*, 166, 333–357.
- Wessells, M. G. (1973). Autosshaping, errorless discrimination, and conditioned inhibition. *Science*, 182, 941–943.
- Williamson, T. (2020). *Suppose and Tell. The Semantics and Heuristics of Conditionals*. Oxford: Oxford University Press.
- Wright, C. (2008). Relativism about truth Itself: Haphazard thoughts about the very idea. In M. García-Carpintero & M. Kölbel (Eds.), *Relative Truth* (pp. 157-85). Oxford: Oxford University Press.

Appendix 1: Formal Details on Semantic Theories

At the core of a semantic theory lies a compositional (recursive) assignment of semantic values to syntactic expressions relative to some collection of parameters $\mathbf{p} = \langle p_1, \dots, p_n \rangle$. We can let $\llbracket e \rrbracket^{\mathbf{p}}$ be the semantic value of the expression e at \mathbf{p} . When sentences are involved the semantic value assigned is typically *true* or *false* (1 or 0). The parameters we use will depend on what kind of syntactic expressions one wishes to capture in the semantic theory. If, for instance, the language contains temporal operators, we need some parameter for time. Most semantic theories insist that there is some parameter denoting the state of the world—a representation of the ‘factual’ properties of the world—and the values for this parameter are usually referred to as *possible worlds*.

Semantic theories for might-modalities typically assess a sentence of the form “It might be the case that A ” (or, simply, “Might A ”) relative to a *set* X of possible worlds, with the idea that Might A is true if A is true at some world in X . Assuming that the only parameters used are a possible world w and a set of worlds X , the standard recursive semantic clause for the might modality then goes:

$$\llbracket \text{Might}A \rrbracket^{w,X} = 1 \text{ if and only if for some } w' \in X: \llbracket A \rrbracket^{w',X} = 1.$$

A basic compositional assignment of semantic values like the one sketched above in and by itself only provides the means to state *semantic* relationships between sentences. For instance, Might($A\&B$) logically entails Might A in virtue of the fact that in every model and for every w and X , if $\llbracket \text{Might}(A\&B) \rrbracket^{w,X} = 1$, then $\llbracket \text{Might}A \rrbracket^{w,X} = 1$. Importantly, the dividing lines between contextualist, relativist and objectivist analyses of might-modalities do not concern such semantic relationships. So, they can in principle agree on the same basic semantic theory. The differences show up when we try to link the basic semantics to norms and conditions of language *use*. This is sometimes referred to as the *postsemantics* and deals with a whole cluster of questions about how to employ the basic semantics when modelling conditions for assertion and assessment of assertions. Here are two such questions:

Assertability Conditions: Under what conditions is a might-sentence properly assertable?

Assessment Conditions: Under what conditions should the assertion of a might-sentence be properly assessed as true?

Consider the question of assertability first. It makes explicit reference to assertability and so implicitly presupposes a speaker, and a context of use c_u (including a potential audience) in which assertions are made. It is typically assumed that the collective or individual beliefs or knowledge—the *information state*—of the participants in c_u can be aggregated into an information state that can be represented as a set of possible worlds X_u . In the simplest case, this set represents what the speaker knows or believes. In more complex accounts, it is a combination of the beliefs and knowledge of the speaker and audience. Let us focus on the simplest case where $X_{c_u} = X_s$ = the set of worlds consistent with what the speaker s believes true. Given a set of possibilities X_s , most accounts would accept some variant of the following:

MightA is *reasonably* assertable for s if and only if for some $w \in X_s$: $\llbracket A \rrbracket^w, X_s = 1$.

That is, it is reasonable to assert MightA if A is consistent with one's information state. The key normative phrase here is *reasonably* assertable, which is here intended as being weaker than *properly* assertable. A competent speaker of English who falsely (but on reasonable good grounds) believes that it is raining in London can *reasonably* assert "It is raining in London", even though the assertion is false. Being reasonable does not free one of culpability. One can, for instance, be expected to apologize for a false assertion and retract it if one finds out that it is false. If a sentence is *properly* assertable, however, there can arise no issue of subsequent culpability or demands for retraction.

For might-sentences the contextualist and objectivist (for different reasons) agree on the stronger normative reading:

MightA is *properly* assertable for s if and only if for some $w \in X_s$: $\llbracket A \rrbracket^w, X_s = 1$.

That is, if A is consistent with the information state then there is nothing wrong *whatsoever* with asserting $\text{Might}A$ (though there may of course be prudential reasons for not asserting $\text{Might}A$). In particular, there is no need to apologize or retract if it turns out that A is false. The relativist, however, does not agree: whether an assertion was *proper*—and so not susceptible to demands of retraction—depends on the information state of who is assessing the claim.

When we turn to conditions for proper assessment we need to take into consideration the information state of the *assessor* a , again represented as a set of possible worlds X_a . The relativist and the objectivist agree on the following:

An assertion of $\text{Might}A$ is *properly* assessed as true if and only if $\llbracket \text{Might}A \rrbracket^{w, X_a} = 1$.

On these accounts one *should* assess a might-claim based on the information state in the context of assessment. The contextualist, however, does not even agree on the normatively weaker claim:

An assertion of $\text{Might}A$ is *reasonably* assessed as true only if $\llbracket \text{Might}A \rrbracket^{w, X_a} = 1$.

That is, on the contextualist account, one does not even come across as *linguistically* competent if one assesses a might-claim relative to one’s own information state.

The differing answers to the questions concerning assertability and assessibility is sufficient to differentiate contextualism, relativism and objectivism.

	Contextualism	Relativism	Objectivism
Assertion of $\text{Might}A$ assessed false if A known false.	No	Yes	Yes
Assertion of $\text{Might}A$ assessed culpable if A is known false.	No	Yes	No

But how do the different accounts arrive at these different answers? Due to different answers to the following questions:

Content Conditions: What does the propositional content of a might sentence depend on?

Truth Conditions: Under what conditions is the proposition expressed by a might-sentence true?

According to the contextualist, the propositional content of a might-sentence depends on the information state of the speaker. If the relevant information state is what the speaker knows, then an assertion of *MightA* has the propositional content “For all I [the speaker] know, *A* is true”.¹⁴ The proposition expressed by an assertion of *MightA* in a context of utterance c_u , in symbols $|\text{MightA}|^{c_u}$, can thus be represented as the set of worlds in which *A* is consistent with the speaker’s beliefs. The proposition $|\text{MightA}|^{c_u}$ is true in w_{c_u} (the world of the context of use), if w_{c_u} is an element of $|\text{MightA}|^{c_u}$.

Given this account of the propositional content of might-claims, this explains why, according to Contextualism, *MightA* is properly assertable if *A* is consistent with the speaker’s beliefs. For $|\text{MightA}|^{c_u}$ is true if *A* is consistent with the speaker’s beliefs. It also explains why it is not even *reasonable* to assess a might-claim relative to the information state of the *assessor*: for the assessor would then be assessing the wrong proposition. In the context of assessment c_a , the proposition expressed by *MightA*, $|\text{MightA}|^{c_a}$ = the set of worlds in which *A* is consistent with the *assessor*’s beliefs. It would be like the following exchange. Jane asserts “My name is Jane” whereby Bill responds by “No you are wrong, my name is Bill”.

The relativist, by contrast, gives a very different account of the propositional content of a might-claim. On this account the content expressed by an assertion of *MightA* does not depend on the information state and so does not differ from speaker to speaker or from speaker to assessor. On the relativist account, an assertion of *MightA* does not attribute any factual property to the world, it does not represent the state of the world in any objective sense. Accordingly, it’s content cannot be represented as a set of possible worlds. The non-

¹⁴ Here a distinction is usually drawn between solipsistic Contextualism, where only the information state of the speaker matters, and the more widespread non-solipsistic version, where the information state of all members of the conversation is decisive (MacFarlane, 2014). On the latter non-solipsistic version of Contextualism, “might *p*” is closer in meaning to “for all we know, *p*”.

representational content of *MightA*, in symbols $|MightA|$, can instead be represented as the set of pairs of worlds and information states (w, X) such that $\llbracket MightA \rrbracket^{w,X} = 1$.

How can such a non-factual proposition be true or false? Here the context of assessment comes in: the truth value of $|MightA|$ depends on the context of assessment, specifically, on the information state of the context of assessment. $|MightA|$ is true in the context of assessment c_a if $|MightA|^{w_{c_a}, X_{c_a}} = 1$. This means that $|MightA|$ can be true relative to one context of assessment and false relative to another; even when they occur in the same world. This explains why, according to Relativism, *MightA* can *reasonably* be asserted if *A* is consistent with the speaker's beliefs. For if the context of assessment is the speaker's own context then $|MightA|$ is true if *A* is consistent with the speaker's beliefs. But it also explains why a speaker who asserts *MightA* can be held culpable—and so be called to retract the assertion—in a context of assessment, where it is known that *A* is false; for in such a context $|MightA|$ is *false*, and an assertion of a false proposition is grounds for culpability.

The objectivist, finally, can take the same view on the propositional content of a might-claim as the relativist: a might-claim does not attribute any property to the world. However, the objectivist deals with this nonrepresentational character of propositional content by shifting focus from the semantic status of the proposition to its epistemic and normative status. The core principle is that a speaker or assessor can *properly* assert or assess *MightA* as true if and only if *A* is consistent with the speaker's or assessor's information state. Whether an assertion or assessment is proper is thus an objective fact of the matter and does not depend on the information state of whoever assesses the propriety of having made the assertion. If one concedes as much then one must also concede that an assertion can be proper yet false. For if *A* is consistent with what a speaker *S* knows, then *S*'s assertion of *MightA* is proper. An assessor who knows that *A* is false will *properly* judge *S*'s assertion as false, but must still acknowledge that *S*'s assertion was proper. With the idea that this combination of proper but false assertions is made possible by the non-representational character of modal content. No

context (of use or assessment) provides any *semantically privileged* information state that determines whether the proposition expressed by the might-claim is true. The truth value of a might-claim is—in the sense in which truth-values form grounds for a normative assessment of an assertion—indeterminate. While one should judge a might-claim as true or false depending on one's information state, there is nothing so semantically special about one's own information state that one can use it as the basis for normative assessment. This means that the *normative* assessment of an assertion (was the assertion proper or should it be retracted as improper?) need not follow the *semantic* assessment of an assertion (was the assertion true or false?).

How one will assess the truth value of a modal assertion will depend on one's information state. While there is no *semantically* privileged information state there is an *epistemically* privileged information state: the state in which one knows all relevant facts. Assuming that it is in principle knowable whether A is true or not, $|\text{Might}A|$ would be assessed true relative to the *best* information state (one in which one knows all relevant facts) if and only if $|A|$ is true. If, as Peirce suggested, *truth lies at the end of enquiry*, this establishes a kind of objective notion of truth for epistemic modals. If w is the actual world then the ideal information state at the end of inquiry is the singleton set $\{w\}$. On this conception of objective truth for epistemic modals, $|\text{Might}A|$ is objectively true in a world w iff $(w, \{w\})$ is an element of $|\text{Might}A|$ if and only if A is true at w . Whether a modal assertion has the property of objective truth or falsity is an entirely objective matter. But this property of objective truth/falsity is to some extent inert: it has neither semantic significance (the conditions of objective truth do not determine semantic content), nor does it enter into conditions of proper assertion (one can *properly*, and so without fear of culpability, assert $\text{Might}A$ even though the assertion is objectively false). Indeed, the only reason for calling the property in question “objective *truth*”, giving it a privileged status, is that a modal proposition will have this property if and only if anyone who knew all the facts would judge it true.

Indeed, once one knows all relevant facts, including, say, that A is false, it makes no sense to say that $|\text{Might}A|$ was true before we learned that A : once one knows all the facts one should judge that it was false all along. It is through this notion of objective truth that Objectivism can explain how convergence in modal knowledge over time is possible.

As Table A1 shows, there are cases which regulate which assertions can be made, and which challenges are appropriate, on which interlocutors can agree, even if they differ on the underlying interpretation of epistemic modals. Thus, despite the uncertainty of whether one is communicating with a person who evaluates might-statements according to Contextualism, Relativism, or Objectivism, it is possible to follow these rules for arguing over might statements and coordinate which challenges need to be addressed by the speaker.

Table A1. Argumentation in a Latent Classes of Different Interpretations of Epistemic Modals without the Norm of Retraction

	Time	Information State	Assertability	Challenges
<i>Case 1</i>	t_1	$i \cap p \neq \emptyset$ p is not excluded by a shared information state.	Might- p is assertable.	1) The speaker is warranted in rejecting challenges to her assertion of might- p . 2) The interlocutor can use factual evidence to challenge the speaker to update the shared information state so that p is excluded and might- p is not assertable.
<i>Case 2</i>	t_2	$i \cap p = \emptyset$ The shared information state has become updated, p is now excluded.	Might- p is not assertable.	1) The interlocutor can warrantably challenge assertions of might- p . 2) The interlocutor can use factual evidence to challenge the speaker to update the shared information state so that p is not excluded and might- p is assertable.
<i>Case 3</i>	t_2	$i \cap p = \emptyset$ The shared information state excludes p at t_2 . At t_1 information accessible to the speaker only, but concealed by the speaker, excluded p .	Might- p is not assertable at t_2 . Might- p was not assertable by the speaker at t_1 .	1) The interlocutor can warrantably challenge assertions of might- p at t_2 . 2) The interlocutor can require that assertions of might- p at t_1 are retracted, because at t_1 might- p was not correctly assertable by the speaker; only the speaker's deception made it appear so.

Note. Case 3 operates with a minimal notion of retraction on which Contextualism, Relativism, and Objectivism can agree even if the more controversial norm of retraction of MacFarlane (2014) does not apply. The difference is that on MacFarlane's norm of retraction, the truth value of might-assertions shifts with the context of evaluations and assertions that were correctly made at t_1 can acquire the status of being in need of retraction if evaluated by a context of assessment with a diverging information state. In contrast, Case 3 deals

with the case in which the might statement was not even appropriately asserted at t_1 according to the speaker's information state, but the speaker concealed this fact to her interlocutors.

Notice moreover that these cases do not rely on the norm of retraction, proposed in MacFarlane (2014) to govern argumentation with epistemic modals. Thus, cases 1-3 in Table A1 illustrate how it is possible to account for argumentation with epistemic modals, even without the norm of retraction.

Appendix 2: Bayesian Hierarchical Latent Trait Model

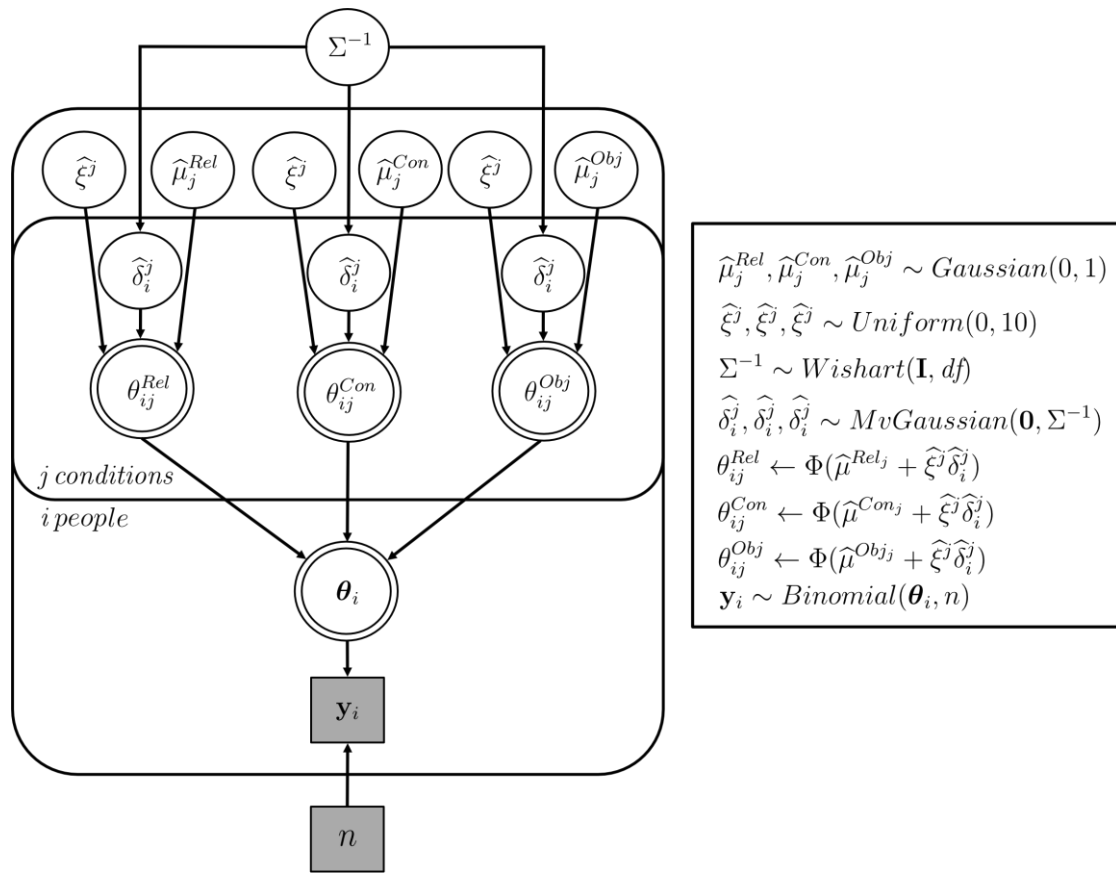
To classify participants into three response styles in a Bayesian analysis, the R-package `rjags` (Plummer, 2019) was used. On this approach, information or ignorance regarding the model parameters is represented by *prior distributions*. The observed data is then used to update our knowledge about the parameters, resulting in *posterior parameter distributions* (Kruschke, 2014; Lee & Wagenmakers, 2014; Skovgaard-Olsen et al. 2019; Skovgaard-Olsen 2019). A Gibbs sampler is used to estimate these posterior distributions by means of Monte Carlo-Markov chains. The general principle is that the posterior distribution of the model parameters is their prior probability times the data likelihood (Li et al., 2018). Accordingly, a Bayesian model is specified in terms of a likelihood function of the data and prior distributions of the parameters. This framework was applied both to estimating the Bayesian latent class models (Tables 10, 11), which were used for assigning class memberships based on participants' performance on the scorekeeping task, and for the hierarchical latent trait model presented below, which was fitted to participants' phase 1 judgments (Table 8). In this way, the classes assigned via the latent class analysis through the scorekeeping task is used to predict individual differences in participants' phase 1 performance (Table 11) by estimating group-specific means in the hierarchical latent trait (Table 1B) analysis as a function of the assigned latent class.

To estimate individual parameters and group-level parameters for the 12 dependent variables in Phase 1 of Experiment 1, the hierarchical latent trait approach of Klauer (2010)

was followed. Instead of aggregating the categorical outcomes across participants, the hierarchical latent trait approach of Klauer (2010) adds a hierarchical structure in which the participants' parameters are constrained to be samples from a population-level probability distribution. On this approach, a probit link function is used to transform MPT parameters (representing probabilities between 0 and 1) to the real line, $\Phi^{-1}(\theta)$. The transformed parameters are then modelled via a multivariate normal distribution while estimating mean, μ , and covariance matrix, Σ , from the data. The advantage of this approach is that heterogeneity in parameter estimates across participants and correlations among personal-level parameters can be accommodated while allowing for partial aggregation of statistical information across participants in the posterior parameters of the multivariate normal distribution (Klauer, 2010). Accordingly, for each participant, i , the probit-transformed parameters are additively decomposed into a group mean, μ , and a random effect, $\Phi^{-1}(\theta) = \mu + \delta_i$.

Instead, of using a multinomial likelihood function for multiple categorical outcomes as Klauer (2010), however, a binomial likelihood function was used to model participants' binary outcomes in Phase 1. Moreover, group-specific means of the hyperparameters were added based on the classification of participants into three latent classes based on the latent class analysis outlined in Table 10.

Table 1B. Hierarchical Latent Trait Binomial Model with Group-Specific Means



Note. There are 12 binary outcome variables. The outcome probabilities of the responses in the data vector, \mathbf{y}_i , are represented by 12 theta parameters for each participant, $\boldsymbol{\theta}_i$, is estimated. There are three group-specific means ($\hat{\mu}^{Rel_j}, \hat{\mu}^{Con_k}, \hat{\mu}^{Obj_j}$) for each of the 12 outcome variables, which are assigned to participants based on which latent class they were classified as in the Phase 2 classification and the latent class analysis in Table 1B. The inverse Wishart distribution has 12+1 degrees of freedom, df , and a 12×12 identity matrix, \mathbf{I} , as a scale matrix.

The models were fitted in a Bayesian framework through the Gibbs sampler implemented in JAGS (Plummer, 2019), which estimates the posterior distributions of model parameters by means of Monte Carlo-Markov chains.