Modality, expected utility, and hypothesis testing

WooJin Chung[1] and Salvador Mascarenhas[2]

[1]Department of Linguistics, Seoul National University.
[2]Institut Jean Nicod, Département d'études cognitives, Ecole Normale Supérieure, EHESS, CNRS, PSL University.

Contributing authors: woojin@snu.ac.kr; salvador.mascarenhas@ens.fr;

Abstract

We introduce an expected-value theory of linguistic modality that makes reference to expected utility and a likelihood-based confirmation measure for deontics and epistemics, respectively. The account is a probabilistic semantics for deontics and epistemics, yet it proposes that deontics and epistemics share a common core modal semantics, as in traditional possible-worlds analysis of modality. We argue that this account is not only theoretically advantageous, but also has far-reaching empirical consequences. In particular, we predict modal versions of reasoning fallacies from the heuristics and biases literature. Additionally, we derive the modal semantics in an entirely transparent manner, as it is based on the compositional semantics of Korean modal expressions that are morphosyntactically decomposed into a conditional and an evaluative predicate.

Keywords: modality, Bayesian confirmation theory, conjunction fallacy, decision theory

1 Introduction

The account of modality due to Angelika Kratzer (1981, 1991, 2012) has been the foundation for many if not most great advances in our understanding of modality in natural language. Over the past decade, this classical account has met challenging objections stemming chiefly from the work of Lassiter (2011, 2017), who proposes an alternative view of epistemic modality grounded in probability measures, and of deontic modality grounded in expected utility. This new perspective on modality has triggered a rich interaction between linguistics and psychology, but not without a cost. Valuable explanatory insights exist in the classical account that find no counterpart in the new approach.

We present an expected value theory of epistemic and deontic modality that preserves one such explanatory insight from the classical theory: all modal expressions share a *core modal semantics*, and their precise modal flavor as epistemic or deontic modals is determined by context. At the same time, our theory shares central properties with Lassiter's account of modality, which proposes that the probability calculus plays a key role in the interpretation of modals. This allows us to explore novel connections between epistemic and deontic semantics and the psychology of probabilistic reasoning, while providing a unified semantics for the two modalities that relies on context to disambiguate modal flavor. Additionally, we provide evidence from Korean modal expressions in support of the particular decomposition of modal semantics we propose. In a nutshell, the prototypical way of expressing modal constructions like English *must* in Korean employs a conditional evaluative. We submit that this evaluative corroborates the expected-value component of our proposal for a semantics for *must*. Finally, our proposal allows for tantalizing connections with a growing literature on Bayesian confirmation-theoretic behavior in human reasoning (Crupi, Elia, Aprà, & Tentori, 2018; Mangiarulo, Pighin, Polonio, & Tentori, 2021; Tentori, Crupi, & Russo, 2013). For the remainder of this introduction, we summarize our proposal, our arguments for it, and its main applications.

Informally, a sentence 'must φ' will be true just in case assuming φ would lead to the only good enough *expected value* among all alternatives to φ, where the calculation of expected value is a function of a contextually supplied body of information. For deontics, expected value will reduce to expected utility. But for epistemics, expected value will be what we call *explanatory value*—an aggregation of the individual probabilities of the propositions in the epistemic background, conditionalized on φ. In this view, epistemic modals do not concern posterior probability of the prejacent, conditional on some epistemic facts. Instead, they assert that the prejacent is the only predictor of contextually relevant epistemic facts which has a good-enough explanatory power. For the simplest case when there is only one contextually relevant epistemic fact, the epistemic reading of 'must φ' against a salient epistemic fact e will reduce to the assertion that only $Pr(e \mid \varphi)$ exceeds the good-enough threshold, whereas every relevant alternative ψ is such that $Pr(e \mid \psi)$ does not meet this standard.

We submit that reconciling the two types of modals is not only theoretically preferable but also has interesting empirical consequences. Our unified theory preserves the decision-theoretic conception of deontic modality via expected utility, as proposed by Lassiter (2011, 2017), allowing us for example to provide an account of the miners puzzle (Kolodny & MacFarlane, 2010).

On the epistemic side, our proposal makes immediate sense of the longstanding intuition that epistemic *must* has a strong evidential flavor. When someone says "It must be raining outside", the hearer typically concludes that that the speaker *inferred* this proposition from some weaker body of evidence, perhaps the fact that someone just entered the room with wet hair. On our view, "It must be raining outside" is true just in case the proposition that it is raining outside offers the only good-enough explanation for a contextually determined, salient body of evidence. Accordingly, we immediately account for the evidential flavor of epistemic *must*.

More tentatively, this view gives us an immediate account of modal variants of reasoning problems from the heuristics and biases literature. For example, in the conjunction fallacy (Tversky & Kahneman, 1983), participants read a description of an individual named Linda that asserts that in her youth she engaged in political activism. Then they are asked to choose which is most likely: (A) Linda is a bank teller, or (B) Linda is a bank teller who is active in the feminist movement. A staggering proportion of participants in the original experiments and countless replications since respond that option (B) is most probable. If participants mean that the probability of (B) conditional on the known facts about Linda is greater than that of (A) conditional on the same facts, they are violating the classical probability calculus. For (B) entails (A), and therefore cannot be more probable than (A) under the same conditionalization. Our theory of modality predicts that participants should be inclined to accept the modal sentence "Linda **must** be a bank teller who is active in the feminist movement" in the same context. The description of Linda constitutes the relevant epistemic background with respect to which the argument of *must* should maximize explanatory value. The sentence will be true only if the probability of the description of Linda conditional on option (B) is greater than the probability of the same description conditional on the alternative (A). Crucially, this assignment of probabilities is by no means incoherent with the probability calculus, and will indeed obtain under any realistic probability distribution. In effect, our theory brings into the realm of modality an account of the conjunction fallacy from psychology that builds on Bayesian confirmation theory (Crupi, Fitelson, & Tentori, 2008; Tentori et al., 2013). Conversely, our theory offers a philosophically-motivated explanation of *why* naive reasoners would opt for inductive reasoning despite fallacious consequences: the deontic counterpart—which uses the same formula to calculate relevant measures but only differs in the body of information attended to—manifests a rational strategy comparing the expected utilities of contextually salient alternatives. What in the deontic domain produces rational behavior by leveraging expected utility, generates a potential for fallacious reasoning in the epistemic domain, by resorting to explanatory value instead of maximizing posterior probabilities.

We derive the modal semantics in an entirely transparent manner. There is linguistic evidence that at least some languages combine conditionals and evaluative predicates to express modal meanings (Ammann & van der Auwera, 2002; Chung, 2019), the compositional semantics of which involves comparing expected utilities (deontic) or confirmation measures (epistemic). Korean is one such language:

(1) Conditional evaluative construction

John-un cip-ey   iss-∅-**eya**   **toy**-n-ta.
John-TOP home-DAT COP-PRES-**only.if** EVAL-PRES-DECL

'(Lit.) Only if John is home, it suffices.'
'John must be home.'

Korean modal expressions are not black boxes in the sense that they are not monomorphemic as in many other languages (e.g., English *must*, *should*, …). These *conditional evaluatives* (M. Kaufmann, 2017) can receive a compositional account

thanks to their transparent morphosyntax. Under the assumption that conditionals roughly denote the degree of support for the consequent given the antecedent (Adams, 1965; Pearl, 2000, 2013), we simply compose our semantics of the evaluative predicate *toy* 'ᴇᴠᴀʟ' with the conditional semantics to derive our theory of modality.

1.1 Extant theories of modality

We briefly introduce two competing theories of modality, one due to Kratzer (1981, 1991, 2012) and the other due to Lassiter (2011, 2017). Our purpose is not to offer a comprehensive review of the two theories, but rather to highlight the notable features of these accounts that ours builds on.

The classical theory due to Angelika Kratzer is a quantification-based approach. The truth conditions of 'must φ' are calculated in two steps: (i) universally quantify over the best worlds and (ii) assert that φ is true in every best world.[1] One of the important insights of the theory is that modal expressions, regardless of their flavor, share a common semantic core. The ambiguity in modal flavor is not due to lexical ambiguity but rather to context sensitivity. Kratzer parameterizes the modal semantics with respect to *conversational backgrounds*, functions from worlds to sets of propositions that are relevant to the interpretation. Each modal is interpreted with respect to a pair of conversational backgrounds. One identifies the set of relevant worlds, and the other is used to pick out the best worlds among the set of relevant worlds. The two conversational backgrounds, the *modal base* and the *ordering source*, jointly identify the domain of quantification of the modal. For epistemics, the modal base represents a set of relevant known facts and the ordering source captures what is stereotypically the case. Accordingly, 'must φ' is true just in case φ stereotypically follows from the relevant known facts. As for deontics, the modal base represents a set of relevant circumstances and the ordering source a set of ideals/goals. 'Must φ' is true just in case φ follows from what is ideally the case given the relevant circumstances.

This context-sensitive analysis of modals nicely captures the crosslinguistic generalization that the majority of modal expressions are ambiguous between an epistemic reading and a deontic reading. We find this context-sensitivity to be an essential feature of any theory of modality.

Lassiter's theory significantly differs from Kratzer's in that the entire theory operates on top of the probability calculus. Lassiter observes that a theory of modality based on a qualitative ordering has difficulties accounting for examples where a degree modifier applied to an epistemic adjective establishes an arithmetic relationship between degrees:[2]

(2)　　a.　It is twice as likely to rain as it is to snow.
　　　　b.　It is 95% certain that it will snow.

---

[1] For clarity of exposition, we make the Limit Assumption (S. Kaufmann, 2017; Lewis, 1973), which states that there exists a set of best accessible worlds for any possible world.

[2] It is important to note that there exist theories that reconcile qualitative-ordering approaches *à la* Kratzer with degree arithmetic (Klecha, 2014; Pasternak, 2016). We thank an anonymous reviewer for pointing this out.

Moreover, Yalcin (2010) has observed that extant theories of comparative modality based on qualitative orderings validate certain normatively invalid modal inferences, like the following:

(3)   a.   Premise 1: φ is as likely as ψ.
       b.   Premise 2: φ is as likely as χ.
       c.   Invalid conclusion: φ is as likely as $(\psi \vee \chi)$.

Lassiter concludes that modal semantics has to encode more quantitative information and builds a theory of modality based on probability distributions. In short, all epistemic necessity modals require that the *probability* of the prejacent be greater than some threshold θ. Weak necessity modals such as *should* or *ought* differ from the strong necessity modal *must* in that θ is sensitive to contextually salient alternatives. As for deontics, weak necessity modals are true just in case the *expected utility* of the prejacent is significantly greater than the contextually-determined threshold θ. The stronger *must* requires a very high θ and also that each of the probable alternatives to the prejacent has an expected utility lower than indifference.

Lassiter's theory has a number of advantages over the classical theory. In particular, the modal inferences it validates are in line with the probability calculus, and it does a better job of explaining the distribution of degree modifiers. However, the innovation comes at the cost of ignoring the cross-linguistic generalization that modals tend to share a common semantic core. In Kratzer's theory, the relevant ordering ranks propositions and has a comparable role to epistemic/deontic measures in Lassiter's theory. The way in which this ordering is calculated does not change depending on the modal flavor. By contrast, there is no single mechanism that derives expected utility and probability in Lassiter's theory. In fact, expected utility is a function of probability, thus the former is a more complex notion than the latter.[3,4]

1.2 The conjunction fallacy and Bayesian confirmation theory

In their seminal 1983 article, Tversky and Kahneman show that human reasoners will often assign subjective probabilities that violate the classical probability calculus in striking ways. The most famous example of this phenomenon is known as the conjunction fallacy, exemplified in (4).

---

[3]Moreover, one cannot reduce the probability weights in an expected utility formula (e.g., (9) on page 11) to the probability of the corresponding proposition. For example, to calculate the expected utility of φ, one needs to consider the probability of each world *conditional on φ*, and use those conditional probabilities as the probability weights of each φ-world. In short, Lassiter's epistemic and deontic measures make use of different kinds of probability, one being an unconditional probability and the other a conditional one.

[4]There have also been attempts to make Kratzer's theory more sensitive to decision-theoretic considerations, e.g., Cariani (2016a), Cariani (2016b), and Blumberg and Hawthorne (2023). Abstracting away from the differences between them, these approaches are sensitive to alternatives just like Lassiter's theory, and rank these alternatives by their expected utilities. Yet, unlike Lassiter's approach, these theories are inherently quantificational in the sense that deontic modals are quantifiers over the best-ranked alternative worlds.

(4)   Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Which is more probable?
a.   Linda is a bank teller.
b.   Linda is a bank teller and she is active in the feminist movement.

Around 85% of participants judged that (4b) was more probable than (4a), and this response was largely independent of the level of education of participants, as well as their field of expertise. However, (4b) entails (4a), and so it must be that $Pr((4b)) \leq Pr((4a))$.

Why bring up the conjunction fallacy in an article about modality? The conjunction fallacy concerns people's intuitions about comparative subjective probabilities, at least *prima facie* it does not seem to involve modality. Yet, there is important connective tissue between modality and comparative subjective probability that we argue makes these facts about reasoning relevant to theories of modality.

First, we observe that both Kratzer's quantification-based theory and Lassiter's probability-based theory relate comparative subjective probabilities to epistemic modality. Concretely, both theories offer accounts of the meanings of words like *must* that are closely related to their accounts of the meanings of words like *probably*. Lassiter's probabilistic theory of modality wears this fact on its sleeve: the meaning of 'must φ' directly appeals to the subjective probability of φ. In Kratzer's account there is no reference to probability *measures*, but the theory provides an account of probability *talk* such as is involved in constructions like 'φ is a better possibility than ψ', and that account is largely shared between constructions like this and *bona fide* modal constructions such as 'must φ'.

Given this theoretical convergence, it is important to ask whether our semantic theories of epistemic and probability operators can shed light on facts about reasoning with epistemic and probability operators.

The theory we present in the next section will do just that, while building on independent tools from formal epistemology. Crupi et al. (2008) provide an account of the conjunction fallacy in (4) in terms of Bayesian confirmation theory. The core idea is that participants in these experiments engage in a kind of hypothesis testing, where (4a) and (4b) are competing hypotheses, and the description of Linda that precedes them is evidence meant to adjudicate between them. Intuitively, (4b) "bank teller active in the feminist movement" is a better theory of the available evidence about Linda than (4a) "bank teller".

There are multiple alternative Bayesian measures of confirmation in the literature (see for example Fitelson 1999), and Crupi et al. (2008) show that all of them work as accounts of the conjunction fallacy. For example, the Difference (D) measure defined below quantifies the extent to which learning some evidence increases one's belief in a particular hypothesis by subtracting the prior from the posterior.

$$D(h, e) = Pr(h \mid e) - Pr(h)$$

Under any plausible probability measure, learning about Linda's prior engagement with various activist movements will increase one's belief in (4b). That is to say, the posterior probability of (4b) conditional on the description is greater than the prior probability of (4b). This is not so for the alternative hypothesis (4a). Sure enough, the posterior probability of (4a) conditional on the description will be higher than that of (4b) conditional on the same description. But crucially the posterior on (4b) *increased more* relative to its prior than the posterior of the alternative (4a) relative to its prior.

An even simpler measure of the explanatory power of a theory can be found in the likelihood of a hypothesis, that is the probability of the evidence conditional on the hypothesis. On this view, hypothesis testing is an intrinsically contrastive task: one should ask "which hypothesis has the greater likelihood for the available evidence?" (Edwards, 1992). As before, any plausible probability measure will ensure that the probability of the description of Linda conditional on (4b) is greater than the probability of the same description conditional on (4a). *Likelihoodism*, as this view is often dubbed, stands in opposition to a multitude of non-contrastive, properly Bayesian measures of hypothesis testing, such as the *D* measure reviewed above (Fitelson, 2007). But even in the Bayesian approach, likelihoods have a role. For example, the likelihood ratio measure *L* below is a respectable Bayesian alternative to the *D* measure, and it will be familiar to any reader acquainted with standard model-comparison techniques say in experimental psychology.

$$L(h, e) = \log\left(\frac{Pr(e \mid h)}{Pr(e \mid \neg h)}\right)$$

A rich literature exists in formal epistemology and philosophy of science on the virtues of the likelihoodist and Bayesian views, and within the latter on the complex trade-offs provided by the various alternative measures of Bayesian confirmation on the market. Our account of modality most straightforwardly produces a likelihoodist view of explanatory adequacy in the epistemic case, as we will show shortly. But we will also illustrate how a more properly Bayesian measure can be achieved.

Before we move on, three important disclaimers are warranted. First, we do not purport here to offer a comprehensive view of the phenomena associated with *reasoning by representativeness*, such as the conjunction fallacy. There is a rich and complex literature on such phenomena that goes beyond the scope of this work. For example, Stolarz-Fantino, Fantino, Zizzo, and Wen (2003) report that the order in which hypotheses are assessed influences people's judgments. Also, there is a closely-related phenomenon dubbed the disjunction fallacy, where people judge a disjunction less probable than its disjunct.[5] There are theories that address these issues, e.g., Busemeyer, Pothos, Franco, and Trueblood's (2011) quantum probability theory assigns

---

[5]While Tentori et al.'s (2013) account of the conjunction fallacy does not extend to the disjunction fallacy, confirmation-theoretic approaches can in principle partially account for it. For example, likelihood-based measures including explanatory value allow for the confirmatory value of a disjunct to be higher than that of a disjunction containing it. Specifically for a likelihoodist view, there is nothing wrong with a probability assignment where $Pr(e \mid h_1) > Pr(e \mid h_1 \lor h_2)$. Nonetheless, there are cases in which likelihood-based approaches, and therefore our own proposal, are not directly applicable: as reported by Morier and Borgida (1984), in the Linda scenario naive reasoners judge that **feminist** ∨ **teller** is more probable than **feminist** ∧ **teller**. This is in line with posterior probabilities and against the predictions of a likelihoodist view.

perceived probabilities to each potential answer to a question under discussion (QUD). Depending on the QUD, the quantum probability theory introduces interference effects which account for the conjunction fallacy, the disjunction fallacy, and the order effects.[6]

Second, in focusing on the benefits of a confirmation-theoretic account of the conjunction fallacy and related phenomena, we do not mean to suggest that such an account explains the entirety of the phenomenon. For example, in the original conjunction-fallacy article, Tversky and Kahneman (1983) consider the possibility that the two options in the Linda problem in (4) are interpreted exclusively. Specifically, option (4a) "bank teller" could be interpreted by contrast with its alternative, and taken to mean "bank teller who is *not* active in the feminist movement". Under such an interpretation, it is no longer a violation of the classical probability calculus to consider option (4a) more probable than option (4b), since one is no longer included in the other. Tversky and Kahneman (1983) control for this pragmatic enrichment in a follow-up experiment, blocking it altogether. They observe that conjunction errors are still prevalent, though their rate dropped from about 85% to about 65%. Later work by Dulany and Hilton (1991) applied a more sophisticated Gricean theory of pragmatics, considering what are now called primary implicatures or ignorance inferences, finding similar results: the conjunction error is mitigated by blocking pragmatic enrichments of the "bank teller" option, but by no means does it disappear. These classical results, replicated multiple times, point to the need for a multi-factor theory of the conjunction fallacy, at least incorporating pragmatic effects. What seems clear is that no single-factor theory of the conjunction fallacy on the market can explain the entirety of the phenomenon. With that said, the confirmation-theoretic view has produced powerful, insightful, and general models of the conjunction fallacy and of other phenomena in the representativeness literature and even deductive reasoning (Sablé-Meyer & Mascarenhas, 2022), demonstrating beyond any doubt its validity as a top contender for an explanation of the non-pragmatic dimension of conjunction errors.

Third and final, we will consider as case studies in this article *modalized* versions of the conjunction fallacy, where the two possible responses to the task to choose from are "Linda **must** be a bank teller" and "Linda **must** be a bank teller and be active in the feminist movement". We will argue based on introspective judgments that such sentences produce conjunction errors much like in the original conjunction fallacy paradigm, and we will show how our account of *must* predicts and explains these putative fallacies. Crucially, we are not prepared to argue that the original conjunction fallacy paradigm ought to be explained in modal terms. That is, we do not propose in this article that silent modals occur in the logical forms of the options in (4), and that those silent modals explain the phenomenon, via our proposed semantics for *must*.[7]

---

[6]While our view does not share many features with the quantum probability theory, it is in line with Busemeyer, Wang, Pothos, and Trueblood's (2015) perspective that theories using inductive confirmation are not incompatible with the quantum probability theory. Rather, one can be used to constrain the other.

[7]We are extremely grateful to an anonymous reviewer for making us see that this possible proposal was compatible and even suggested in an earlier draft of this article. See also Footnote 21 for additional comments about this intriguing theoretical possibility.

2 Proposal

We propose that necessity modals compare the probability-weighted measure of the prejacent to the probability-weighted measure of each of its alternatives. Specifically, 'must φ' is true if and only if the expected value of the prejacent is (significantly) greater than the contextually determined threshold, but the expected value of each alternative to φ does not exceed the threshold. Depending on the flavor of the modal, expected value either corresponds to expected utility or explanatory value. The flavor is determined by a single parameter R, which represents a set of ideals/rules for deontics and a set of relevant known facts in need of an explanation (i.e., pieces of evidence) for epistemics. Alternatives to the prejacent φ in our proposal are determined by the context in the shape of a question under discussion if available: for the deontic case, a set of possible courses of action under consideration and for the epistemic case, a set of candidate explanations for the salient body of information at hand.[8]

To formalize our proposal, we first define $\mathbb{E}[\psi \mid \varphi]$ as in (5).[9] It is the probability-weighted average of the value of ψ over φ-worlds normalized with respect to the probability of φ. This is equivalent to the expected value of ψ conditioned on φ. We parameterize the probability function $Pr(\cdot)$ with respect to the world of evaluation—accordingly the expected value function $\mathbb{E}$ as well—to reflect that probability assignments are world dependent.

(5) $\mathbb{E}_w[\psi \mid \varphi]$ evaluated at w

$$\mathbb{E}_w[\psi \mid \varphi] = \frac{1}{Pr_w(\varphi)} \sum_{w_j \in \varphi} \psi(w_j) \times Pr_w(\{w_j\})$$

$$= \sum_{w_j \in \varphi} \psi(w_j) \times \frac{Pr_w(\{w_j\})}{Pr_w(\varphi)}$$

$$= \sum_{w_j \in \varphi} \psi(w_j) \times Pr_w(\{w_j\} \mid \varphi)$$

We will later elaborate on how this relates to expected utility or explanatory value. Also, we will show in section 4 that the compositional semantics of Korean conditional evaluatives serves as natural language evidence that at least some modals employ the above expected-value calculation.

Our formal analysis of modal necessity is given in (6), which reads as follows: For deontics, the expected utility of φ is greater than θ but no alternative to φ is such that its expected utility is greater than θ.[10] For epistemics, the explanatory value of φ

---

[8]We will have little more to say about sources of alternatives in this article. In the absence of a rich-enough context to determine alternatives, we assume that the presence of focus in a constituent of the prejacent will trigger a QUD and set of alternatives as in the classical analysis of the semantics of focus, and that a simple set of polar alternatives based on the prejacent constitutes a default fallback mechanism. These considerations are in line with related instances of sensitivity to alternatives in the modal domain (Dretske, 1972; Heim, 1992; Villalta, 2008)

[9]For convenience, we use Greek letters to represent both object- and meta-language formulae.

[10]Cariani (2016a) convincingly shows that theories of expected value contrastivism, along with actualist theories such as Jackson (1985) and Jackson and Pargetter (1986), invalidate the plausible inference of *Weakening*: ought(φ), ought(ψ) ⊨ ought($\varphi \lor \psi$). Our theory is in the same situation, for it contrasts the expected utilities of salient alternatives. We thank an anonymous reviewer for pointing this out.

is greater than θ but no alternative to φ is such that its explanatory value is greater than θ. We use the notation $Alt(\varphi)$ to indicate the set of alternatives to φ, abstracting away from the details of how they are determined.

(6)   Proposal

$$[\![\text{ must } \varphi \;]\!]^w = (\mathbb{E}_w[\mu_{\text{EVAL}} \mid \varphi] > \theta) \wedge \forall \psi \in Alt(\varphi) : (\mathbb{E}_w[\mu_{\text{EVAL}} \mid \psi] \leq \theta)$$

We find it useful and intuitive to read the formula as follows: In a deontic context, φ is *the only good-enough choice* among the available options. In an epistemic context, φ is *the only good-enough explanation* of the evidence among the available hypotheses.

We define μ_{EVAL} as a measure function which takes a world argument and returns the degree to which the given world supports the contextually-supplied body of information R. Technically, this amounts to counting the number of relevant propositions $r \in R$ that are true at w.

(7)   Measure function over worlds

$$\mu_{\text{EVAL}} = \lambda w. \mid \{r \in R \mid r \text{ is true at } w\} \mid,$$
$$\text{where } R \text{ is the set of relevant propositions}$$

As in Kratzer's standard theory, a single parameter determines the flavor of a modal. Conversational backgrounds determine the flavor in Kratzer's theory, and R—a set of relevant propositions—in ours.

Let us first demonstrate how $\mathbb{E}_w[\mu_{\text{EVAL}} \mid \varphi]$ yields the expected utility of φ in the deontic case. For deontics, the measure function employs a deontic R_D, which characterizes the set of relevant rules or ideals. The measure function μ_{EVAL} takes a world w and checks how many ideals/rules $d \in R_D$ are realized/abided by at w (technically, true at w). The more ideals/rules are realized/abided by at w, the better the world w is. In this sense, the number of ideals/rules realized/abided by at a given world is the *utility value* of the world. Thus, we can interpret μ_{EVAL} as a function which takes a world and returns the utility value of the world argument.

(8)   Deontic interpretation of μ_{EVAL}

$$\mu_{\text{EVAL}} = \lambda w. \mid \{d \in R_D \mid d \text{ is true at } w\} \mid,$$
$$\text{where } R_D \text{ is the set of relevant ideals/rules}$$

Replacing ψ with μ_{EVAL} in (5) yields the following, which demonstrates that $\mathbb{E}[\mu_{\text{EVAL}} \mid \varphi]$ corresponds to the expected utility of φ:

10

(9)   Deontic measure: $\mathbb{E}_w[\mu_{\text{EVAL}} \mid \varphi]$ as the expected utility of φ at w

$$\mathbb{E}_w[\mu_{\text{EVAL}} \mid \varphi] = \frac{1}{Pr_w(\varphi)} \sum_{w_j \in \varphi} \mu_{\text{EVAL}}(w_j) \times Pr_w(\{w_j\})$$

$$= \sum_{w_j \in \varphi} \mu_{\text{EVAL}}(w_j) \times \frac{Pr_w(\{w_j\})}{Pr_w(\varphi)}$$

$$= \sum_{w_j \in \varphi} \mu_{\text{EVAL}}(w_j) \times Pr_w(\{w_j\} \mid \varphi)$$

The formula conditionalizes on φ, and for each φ-world, it calculates the utility value of the world. It then calculates the probability-weighted average of the utility values of φ-worlds. This is by definition the expected utility of φ.[11]

Let us turn to the epistemic case. The epistemic interpretation of μ_{EVAL} employs an epistemic R_E, which characterizes the set of relevant known facts (i.e., pieces of evidence).

(10)   Epistemic interpretation of μ_{EVAL}

$$\mu_{\text{EVAL}} = \lambda w. \mid \{e \in R_E \mid e \text{ is true at } w\} \mid,$$
where R_E is the set of relevant known facts

For the epistemic interpretation of $\mathbb{E}_w[\mu_{\text{EVAL}} \mid \varphi]$, we find it more intuitive to reformulate the measure function μ_{EVAL} as in (11). The two formulae are equivalent since each $e \in R_E$ is a proposition (i.e., returns 1 if true and 0 otherwise). Using this formulation, (12) shows that $\mathbb{E}_w[\mu_{\text{EVAL}} \mid \varphi]$ denotes the sum over the probabilities of each relevant known fact $e_i \in R_E$ conditionalized on φ. In other words, it is the sum over the likelihoods (i.e., inverse probabilities) of φ with respect to each relevant known fact $e_i \in R_E$.[12]

(11)   Reformulation of μ_{EVAL} interpreted with respect to R_E

$$\mu_{\text{EVAL}} = \lambda w. \mid \{e \in R_E \mid e \text{ is true at } w\} \mid$$

$$= \lambda w. \sum_{i=1}^{n} e_i(w), \quad \text{where } R_E = \{e_1, ..., e_n\}$$

---

[11] Along with Jeffrey (1965) and Gibbard and Harper (1978), we represent outcomes in terms of possible worlds without introducing a separate ontology of outcomes.

[12] An anonymous reviewer notes that this analysis crucially assumes that the cardinality of R_E is finite, and asks what would happen if the cardinality of R_E was infinite. We agree that things get tricky when there are infinite pieces of evidence. However, given that R_E represents a salient body of evidence to be explained—something one should be able to entertain in their mind at the time of utterance—we think that this is a reasonable assumption, although we acknowledge this is yet another departure from the standard semantics.

(12)   Epistemic measure: $\mathbb{E}_w[\mu_{\text{EVAL}} \mid \varphi]$ as the explanatory value of φ at w

$$\mathbb{E}_w[\mu_{\text{EVAL}} \mid \varphi] = \sum_{w_j \in \varphi} \mu_{\text{EVAL}}(w_j) \times Pr_w(\{w_j\} \mid \varphi)$$

$$= \sum_{w_j \in \varphi} \sum_{i=1}^{n} e_i(w_j) \times Pr_w(\{w_j\} \mid \varphi)$$

$$= \sum_{i=1}^{n} \sum_{w_j \in \varphi} e_i(w_j) \times Pr_w(\{w_j\} \mid \varphi)$$

$$= \sum_{i=1}^{n} Pr_w(e_i \mid \varphi), \quad \text{where } R_E = \{e_1, ..., e_n\}$$

In the simplest case where there is only one piece of evidence, say e, the expected value of φ reduces to the likelihood of φ with respect to e at w. Since this is one way to cash out the degree to which evidence e supports and is explained by φ, we call this measure the *explanatory value* of φ.

This analysis of epistemic modality is sharply different from Lassiter's. Lassiter argues that epistemic modals compare the *(posterior)* probability of the prejacent to a contextually determined threshold, whereas we propose that epistemic modals are concerned with the *explanatory value* of φ which is based on likelihoods.

Note that the proposed semantics *indirectly* compares the expected value of the prejacent to those of its alternatives: it conveys that the expected value of φ is greater than those of its alternatives by asserting that only the former is greater than θ. There is an alternative formulation (though not equivalent) that makes *direct* comparisons, and that under certain conditions produces the L confirmation measure mentioned in section 1.2. The alternative formulation in (13) conveys that the expected value of φ is greater than the expected values of its alternatives by at least θ.

(13)   Alternative analysis

$$[\![\text{must}_{\text{ALT}} \, \varphi \,]\!]^w = \forall \psi \in Alt(\varphi) : \mathbb{E}_w[\mu_{\text{EVAL}} \mid \varphi] - \mathbb{E}_w[\mu_{\text{EVAL}} \mid \psi] > \theta$$

If we assume that (i) the only alternative to φ is its negation, (ii) there is a single piece of evidence,[13] and (iii) take the logarithm of each measured value, 'must φ' is true if and only if $L(\varphi, e)$ is greater than the contextually supplied threshold θ, as shown below:[14]

---

[13]In fact, it is common practice in the confirmation-theoretic literature on the conjunction fallacy for example to conjoin several pieces of evidence into a single proposition. In our terms, this amounts to conjoining the relevant known facts $e_i \in R_E$ and using the conjunction as evidence. It is an open question whether this is a necessary move, and there will be differences in empirical predictions depending on whether one considers the conjunction of R_E or the set with multiple pieces of evidence.

[14]The purpose of using logarithms is to interpret positive values as confirmation, zero as irrelevance, and negative values as disconfirmation. Therefore, our alternative formulation of modality can be understood as directly encoding the L confirmation measure.

(14)   Log-based alternative analysis (assuming $Alt(\varphi) = \{\neg\varphi\}$)

$$
\begin{aligned}
[\![\text{must}_{\text{LOG,ALT}} \; \varphi \;]\!]^w &= \forall \psi \in Alt(\varphi) : log(\mathbb{E}_w[\mu_{\text{EVAL}} \mid \varphi]) - log(\mathbb{E}_w[\mu_{\text{EVAL}} \mid \psi]) > \theta \\
&= log(\mathbb{E}_w[\mu_{\text{EVAL}} \mid \varphi]) - log(\mathbb{E}_w[\mu_{\text{EVAL}} \mid \neg\varphi]) > \theta \\
&= log(Pr_w(e \mid \varphi)) - log(Pr_w(e \mid \neg\varphi)) > \theta \\
&= log\left(\frac{Pr_w(e \mid \varphi)}{Pr_w(e \mid \neg\varphi)} \right) > \theta \\
&= L(\varphi, e) > \theta
\end{aligned}
$$

Both our proposal in (6) and the alternative in (13) are enough to capture the evidential flavor of typical utterances involving epistemic *must*. Imagine someone conspicuously enters the room soaking wet. In so doing, they establish a set R_E of salient information in need of an explanation, say simply the singleton set containing a proposition to the effect that "This person is wet". On our account, an onlooker might now utter "It must be raining", only if rain is the only good-enough explanation for the salient body of evidence at hand, as is intuitively the case.

An interesting implication arises from our theory of modality: people's conception of modality facilitates rational decision making with deontics, but the very same mechanism can be a source of irrationality when assessing comparative subjective probabilities with epistemics. Note that expected utility is a rational measure employed in decision theory. By contrast, explanatory value in terms of confirmation theory is a measure that will often diverge from the standard of probabilistic rationality offered by posterior probabilities, which form the basis of all other extant probabilistic accounts of *must*. Our theory, then, predicts an undersized role for rational posterior probabilities in epistemic utterances with *must*.

3 Case studies

We present three case studies that our theory accounts for and explains. We start with the miners puzzle on the deontic side (Kolodny & MacFarlane, 2010). For epistemics, we discuss two related but distinct examples from the heuristics and biases literature: the conjunction fallacy and base-rate neglect (Kahneman & Tversky, 1973; Tversky & Kahneman, 1983).

3.1 The miners puzzle (Kolodny & MacFarlane, 2010)

As Lassiter (2011) points out, an expected-utility theory of deontic modality naturally addresses the issue of interpreting modals under epistemic uncertainty. A representative case of the issue is known as the miners puzzle, given in (15) and summarized in Table 1 (Kolodny & MacFarlane, 2010). Given the situation described in Table 1, examples (15a)–(15c) are all intuitively true.

(15)   Ten miners are trapped either in shaft A or in shaft B, but we do not know which. Flood waters threaten to flood the shafts. We have enough sandbags to block one shaft, but not both. If we block one shaft, all the water will go

Table 1 Summary of possible outcomes in the miners puzzle, following Kolodny and MacFarlane (2010)

Action	If miners in A	If miners in B
Block shaft A	All saved	All drowned
Block shaft B	All drowned	All saved
Block neither shaft	One drowned	One drowned

into the other shaft, killing any miners inside it. If we block neither shaft, both shafts will fill halfway with water, and just one miner, the lowest in the shaft, will be killed.

    a.   We ought to block neither shaft.
    b.   If the miners are in shaft A, we ought to block shaft A.
    c.   If the miners are in shaft B, we ought to block shaft B.

However, the classical theory of modality predicts that the three examples cannot all be true. Below is a proof sketch:[15]

(16)   a.   'We ought to block neither shaft' is true if and only if among the circumstantially relevant worlds (say R_1), the best worlds in R_1 are such that 'we block neither shaft' is true.

    b.   'If the miners are in shaft A, we ought to block shaft A' is true if and only if (i) the circumstantially relevant worlds are restricted to worlds where the miners are in shaft A (say R_2) and (ii) the best worlds in R_2 are such that 'we block shaft A' is true.

    c.   'If the miners are in shaft B, we ought to block shaft B' is true if and only if (i) the circumstantially relevant worlds are restricted to worlds where the miners are in shaft B (say R_3) and (ii) the best worlds in R_3 are such that 'we block shaft B' is true.

    d.   If any best world in R_1 (say w_1) is a member of R_2, then w_1 is also a best world in R_2. This implies that (15a) and (15b) cannot both be true because 'we block neither' and 'we block shaft A' cannot both be true at w_1.

    e.   If any best world in R_1 (say w_2) is a member of R_3, then w_2 is also a best world in R_3. This implies that (15a) and (15c) cannot both be true because 'we block neither' and 'we block shaft B' cannot both be true at w_2.

    f.   Any best world in R_1 is either a member of R_2 or R_3 because $R_1 = R_2 \cup R_3$ (i.e., either the miners are in shaft A or shaft B). Then from (16d) and (16e), examples (15a)-(15c) cannot all be true.

Kolodny and MacFarlane argue that the issue arises because Kratzer's conversational backgrounds are not seriously information-dependent, that is, one's preferences cannot change upon obtaining new information.

An expected-utility analysis of the miners puzzle naturally encodes this information dependence into the semantics, as conditionalizing on new information adjusts

---

[15]For details concerning other possible interpretations, we refer the reader to Kolodny and MacFarlane (2010) and Cariani, Kaufmann, and Kaufmann (2013).

the probability weights used to calculate expected utilities (Lassiter, 2011). Our common-core semantics for modality in terms of expected value reduces to expected utility in the deontic case, as we explained above. This means that our approach should be able to resolve the miners puzzle without much difficulty. In what follows, we show that this is the case.

First, notice that the miners puzzle as phrased in the literature and in (15) above is a puzzle about *ought*, rather than *must*. We will address and assess our predictions for *must*-sentences in this scenario at the end of this section. For now, we give the simplest possible semantics for *ought* (and *should*, for that matter) that keeps with the spirit of our proposal for *must* in this article. Specifically, we propose that 'ought φ' is true just in case φ is the best good-enough option among the alternatives under consideration.

(17)   Proposal for *ought*

$\llbracket$ ought φ $\rrbracket^w = (\mathbb{E}_w[\mu_{\text{EVAL}} \mid \varphi] > \theta) \wedge \forall \psi \in Alt(\varphi) : (\mathbb{E}_w[\mu_{\text{EVAL}} \mid \varphi] > \mathbb{E}_w[\mu_{\text{EVAL}} \mid \psi])$

This is the semantics for 'must φ', minus the requirement that φ be *the only* good-enough alternative. This simple approach is motivated by observations very much in this direction in the literature on teleological modality (von Fintel & Iatridou, 2005), and on work specifically on weak necessity modals (Finlay, 2009; Goble, 1996; Jackson, 1985; Sloman, 1970).

Regarding 'we ought to block neither shaft', in this analysis the requirement is that the expected utility of blocking neither shaft (i.e., **block-neither**) is higher than the contextual threshold θ, and greater than the expected utility of blocking shaft A (i.e., **block-A**) and the expected utility of blocking shaft B (i.e., **block-B**). We posit the following R_D, which was borrowed from Cariani et al. (2013):

(18) $R_D = \{$ 1 miner saved, 2 miners saved, . . . , 10 miners saved $\}$

We take it that the subjective probabilities of the miners being in shaft A, respectively shaft B, are both 0.5. Given these background assumptions, μ_{EVAL} returns 9 as the utility for each **block-neither**-world. This is because the context guarantees that 9 miners will be saved if we block neither shaft. Consequently, the expected utility of **block-neither** is 9, as we show in (19).

(19)   The expected utility of 'we block neither shaft'

$$\mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{block-neither}]$$
$$= \sum_{w_j \in \textbf{block-neither}} \mu_{\text{EVAL}}(w_j) \times Pr_w(\{w_j\} \mid \textbf{block-neither})$$
$$= \sum_{w_j \in \textbf{block-neither}} 9 \times Pr_w(\{w_j\} \mid \textbf{block-neither})$$
$$= 9 \times \sum_{w_j \in \textbf{block-neither}} Pr_w(\{w_j\} \mid \textbf{block-neither})$$
$$= 9$$

15

On the other hand, μ_{EVAL} returns 10 for each **block-A** $\wedge$ **miners-in-A**-world, and 0 for each **block-A** $\wedge$ **miners-in-B**-world. As we show in (20), the expected utility of **block-A** is 5 assuming that **miners-in-A** and **miners-in-B** are equally probable and the propositions representing our actions and the miners' whereabouts are independent. Analogously, the expected utility of **block-B** is also 5.

(20)   The expected utility of 'we block shaft A'

$$
\begin{aligned}
&\mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{block-A}] \\
&= \sum_{w_i \in \textbf{block-A}} \mu_{\text{EVAL}}(w_i) \times Pr_w(\{w_i\} \mid \textbf{block-A}) \\
&= \sum_{w_j \in \textbf{block-A} \wedge \textbf{miners-in-A}} \mu_{\text{EVAL}}(w_j) \times Pr_w(\{w_j\} \mid \textbf{block-A}) \\
&\qquad + \sum_{w_k \in \textbf{block-A} \wedge \textbf{miners-in-B}} \mu_{\text{EVAL}}(w_k) \times Pr_w(\{w_k\} \mid \textbf{block-A}) \\
&= \sum_{w_j \in \textbf{block-A} \wedge \textbf{miners-in-A}} 10 \times Pr_w(\{w_j\} \mid \textbf{block-A}) \\
&\qquad + \sum_{w_k \in \textbf{block-A} \wedge \textbf{miners-in-B}} 0 \times Pr_w(\{w_k\} \mid \textbf{block-A}) \\
&= 10 \times \sum_{w_j \in \textbf{block-A} \wedge \textbf{miners-in-A}} Pr_w(\{w_j\} \mid \textbf{block-A}) \\
&= 10 \times Pr_w(\textbf{block-A} \wedge \textbf{miners-in-A} \mid \textbf{block-A}) \\
&= 10 \times Pr_w(\textbf{miners-in-A} \mid \textbf{block-A}) \\
&= 10 \times Pr_w(\textbf{miners-in-A}) \\
&= 10 \times 0.5 = 5
\end{aligned}
$$

(21)   The expected utility of 'we block shaft B'

$$
\mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{block-B}] = 5
$$

We analyze (15a) as in (22). Informally, "blocking neither shaft is the best good-enough choice among the available options". The sentence is accurately predicted to be true, under the reasonable assumptions we've been making about the probability distribution underlying this scenario.

(22)   Analysis of 'we ought to block neither shaft'

$$
\begin{aligned}
[\![\text{ought } \textbf{block-neither}]\!]^w = {}&(\mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{block-neither}] > \theta) \\
&\wedge (\mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{block-neither}] > \mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{block-A}]) \\
&\wedge (\mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{block-neither}] > \mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{block-B}])
\end{aligned}
$$

We turn to the analysis of the deontic conditional in (15b). Following Lassiter (2011), we take it that the *if*-clause requires the expected utility calculation to additionally conditionalize on the antecedent proposition.[16]

Conditionalizing on **miners-in-A** does not change the expected utility of **block-neither** since exactly one miner will drown irrespective of the location of the miners. However, this does raise the expected utility of **block-A**, as we show in (23). The expected utility of **block-A**, assuming **miners-in-A**, is 10, which is greater than 9, the expected utility of **block-neither**. Moreover, the conditionalization on **miners-in-A** reduces the expected utility of **block-B** to 0. The upshot is that the expected utility of **block-A** is greater than the expected utilities of **block-neither** and **block-B**.

(23)   The expected utility of 'we block shaft A' conditionalized on 'the miners are in shaft A'

$$\mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{miners-in-A} \wedge \textbf{block-A}]$$

$$= \sum_{w_i \in \textbf{miners-in-A} \wedge \textbf{block-A}} \mu_{\text{EVAL}}(w_i) \times Pr_w(\{w_i\} \mid \textbf{miners-in-A} \wedge \textbf{block-A})$$

$$= \sum_{w_i \in \textbf{miners-in-A} \wedge \textbf{block-A}} 10 \times Pr_w(\{w_i\} \mid \textbf{miners-in-A} \wedge \textbf{block-A}) = 10$$

We flesh out our analysis of (15b) in (24). Informally, "given that the miners are in shaft A, blocking shaft A is the best good-enough choice among the available options".

(24)   Analysis of 'If the miners are in shaft A, we ought to block shaft A'

$$[\![\text{ if } \textbf{miners-in-A}, \text{ ought } \textbf{block-A}]\!]^w$$
$$= \mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{miners-in-A} \wedge \textbf{block-A}] > \theta$$
$$\wedge\, (\mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{miners-in-A} \wedge \textbf{block-A}]$$
$$> \mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{miners-in-A} \wedge \textbf{block-neither}])$$
$$\wedge\, (\mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{miners-in-A} \wedge \textbf{block-A}]$$
$$> \mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{miners-in-A} \wedge \textbf{block-B}])$$

What we presented in this section is more or less a reproduction of Lassiter's analysis.[17]

---

[16]There is independent motivation for this assumption. In section 4, we show that the expected utility of φ can be derived from the compositional semantics of 'if φ, then EVAL/suffice', which we claim to be part of the underlying logical representation of modal necessity. Under this hypothesis, the analysis of 'if **miners-in-A**, ought **block-A**' involves interpreting 'if **miners-in-A**, then if **block-A**, then EVAL/suffice', which is equivalent to 'if **miners-in-A** ∧ **block-A**, then EVAL/suffice' if we take the Import-Export Principle (Gibbard, 1980; McGee, 1985) for granted. Given the assumptions to be presented in section 4, the latter denotes the expected utility of **miners-in-A** ∧ **block-A**.

[17]Precisely speaking, Lassiter's semantics for *ought* does not directly compare the expected utility of the prejacent to its alternatives, but rather to a contextually-determined threshold. There is also some difference between our analysis of deontic *must* and Lassiter's: while both theories submit that the expected utilities of the alternatives are somewhat low, the latter imposes a stronger requirement, namely that they are lower than the expected utility of indifference—the union of salient alternatives. Lassiter thus would make the wrong prediction that (25b) and (25c) are false, because although **block-neither** is not the best choice given that the miners are in shaft A (outranked by **block-A**), its expected utility is still much higher than that of indifference, i.e., **block-neither** ∪ **block-A** ∪ **block-B**.

This is no surprise because both theories compare expected utilities of contextually salient alternatives. Things start becoming more interesting, in our view, once we consider the predictions for the strong necessity modal *must*. Sticking to the same scenario as described in (15), consider now the following sentences:

(25) a. We must/have to block neither path.
      b. If the miners are in shaft A, we must / have to block shaft A.
      c. If the miners are in shaft B, we must / have to block shaft B.

We submit first of all that (25b) and (25c) are just as felicitous, and crucially ring just as true, as their *ought* variants. Our judgments are less sharp for (25a), but we suspect that an alternative reading, with 'neither' scoping above the modal, is causing interference. Notice that we can rephrase (25a) to unambiguously zoom in on the intended reading:

(26) a. We mustn't block either path.
      b. We must/have to refrain from blocking either path.
      c. We cannot block either path.

We will address possibility modals as in (26c), in Section 6. For now, we take it that (26a) and (26b) are felicitous and true in the scenario at hand.

With our semantics for *must*, (26a) and (26b) will be true just in case **block-neither** is *the only* good-enough option, a stronger set of truth conditions than for the *ought* variant. These truth conditions will still obtain very easily: recall that the expected utility for **block-neither** is 9, while that of each of its alternatives **block-A** and **block-B** is 5. It will therefore be trivial to find a threshold θ between 5 and 9 to ensure that the sentence is true.

The situation is more complex for the conditional sentences in (25b) and (25c). Take (25b), without loss of generality. We predict that this sentence will be true just in case **block-A** is *the only* good-enough option, once we assume that the miners are in shaft A. Now, as we showed above, the expected utility of **block-A** in this conditionalization is 10, while the expected utility of **block-neither** is 9 and that of **block-B** is 0. But if *must* requires that the prejacent be *the only* alternative above the threshold, then we will need for our threshold to be $10 > \theta \geq 9$, while for the unconditional sentence in (26a) we had that $9 > \theta \geq 5$. These two requirements are of course incompatible.

This intriguing tension, in that a shift of standards of evaluation θ is required to judge all the sentences in (25) as true, will emerge not only in our analysis of a *must* variant of the miners puzzle, but indeed in any account of the original puzzle that requires that the utility of the prejacent at hand be *the only* one above the standard of evaluation θ. Such an account for example is sketched by Lassiter (2011): Suppose that there are good reasons to spend my vacation with my parents whom I have not met for a long time and, although incompatible with the first plan, visit my ailing grandparents. Lassiter notes that both of the following sentences are odd because there isn't a unique best option with significant probability which is better than being indifferent:

(27) a. I must go to my parents.
      b. I must go to my grandparents.

We cannot fully resolve the issue in this article, but we have two remarks we think are promising. First, the idea of shifting thresholds so easily might actually not be much of a problem. It is plausible for thresholds of this sort to be highly sensitive to the set of alternatives under consideration and to the modal base in question. Regarding alternatives, it seems clear that deontic *must* sentences will be felicitous with prejacents that are quite "bad", so long as the fully transparent alternative set is exhaustive with respect to all plausible possibilities and has the property that none of the alternatives is "good" in a positive or absolute sense. We conjecture further that thresholds might be able to shift seamlessly depending on different modal bases, that is in our terms different conditionalizations, as is the case in the threshold tension at hand with the miners puzzle. Additionally, an interesting fine-grained prediction emerges from this need for shifting thresholds, shared by any account of the relevant operators that requires that the prejacent be *the only* good-enough alternative. We predict that there should be some processing signature of the shift in θ between judging the truth of the unconditional sentence and the truth of the conditional sentences.[18]

3.2 The conjunction fallacy (Tversky & Kahneman, 1983)

Recall the most well-known variant of the conjunction fallacy, accepted by about 85% of experimental subjects (Tversky & Kahneman, 1983).

(28) Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable?
      a. Linda is a bank teller.
      b. Linda is a bank teller and she is active in the feminist movement.

As we argued in the introduction, there is a convergence between both Kratzer and Lassiter's theories of modality regarding the connection between epistemic modality and probability talk. This theoretical convergence at the very least primes the question whether we find with *must* the same reasoning behavior that we find with *probable*. Specifically for the conjunction fallacy, we propose that a large proportion of experimental subjects would commit a *modal* conjunction fallacy: when faced with the same setup as the classical task, people would generally find (29b) a more attractive response than (29a).[19]

---

[18]Bouletics display a similar sort of threshold-shifting effect (Blumberg & Hawthorne, 2022; Crnič, 2011; Lassiter, 2011). An anonymous reviewer brings up the following case: suppose you like pizza much more than any of the other options on the menu, and you like ramen only a little more than hotdogs. In this context, "I want pizza but if they don't have pizza, then I want ramen" sounds true. If desire verbs were to be given scalar semantics as in Lassiter (2011), the felicity of the aforementioned example can only be explained in terms of threshold-shifting under conditionalization.

[19]An anonymous reviewer points out that the *must* version in (29b) sounds odd to their ears, while a version with *ought* substituted for *must* is appreciably more felicitous. One very plausible source for this judgment is the strong semantics of *must*, an idea somewhat reminiscent of Lassiter's (2011) analysis of deontic *must*: the prejacent is *the only* good-enough explanation for the salient body of evidence. Indeed, Linda's being "a bank teller who is active in the feminist movement"

(29) a. Linda must be a bank teller.
      b. Linda must be a bank teller and be active in the feminist movement.

The original conjunction fallacy asked participants to pick the option that was most probable, but this task becomes somewhat odd when the options to choose from are modal statements as in (29).[20] The roots of the oddity are unclear. In theories where modal operators involve conditions on probabilities, such as ours, this task would require a judgment of the probability of a certain statement about probabilities, which is by no means incoherent, as consistent theories of higher-order probabilistic statements exist (Gaifman, 1988). But it is an unusual move, and one where there is no consensus on what the right theory is, so that it is best to avoid this and other complications arising from embeddings of probability and modality (Goldstein & Santorio, 2021).

We propose to evaluate our prediction in a betting paradigm. In one of their experiments, Tversky and Kahneman (1983) asked participants to bet on one of the statements about Linda. They observed some mitigation of conjunction errors, a drop from about 85% to about 65% error rates. While the reason for this mitigating effect of the betting paradigm is unclear, the result is still that sizable conjunction errors were observed. Applying this paradigm to our proposed modal conjunction fallacy, the task would be to decide on one of the two modal statements in (29) to bet on, thus avoiding the linguistic and conceptual awkwardness of explicitly attempting to assess the probability of a modal statement.

To the best of our knowledge, the heuristics and biases literature, or the modality literature for that matter, has not investigated this issue experimentally. Yet introspection tells us and a group of informants in our social circles that (29b) is in a clear sense more attractive than (29a). Introspection is an entirely valid means of establishing empirical facts under the appropriate circumstances, and we submit that those conditions obtain in the case at hand.

For concreteness, we provide reasonable probability assignments concerning the Linda scenario in (30) and (31). We restricted our attention to the two most relevant pieces of information about Linda, namely that she was deeply concerned with issues of discrimination and social justice (i.e., **social-justice**) and participated in anti-nuclear demonstrations (i.e., **anti-nuclear-protests**).

(30) $Pr_w(\textbf{social-justice} \mid \textbf{teller}) = 0.3$
      $Pr_w(\textbf{anti-nuclear-protests} \mid \textbf{teller}) = 0.2$

(31) $Pr_w(\textbf{social-justice} \mid \textbf{feminist-teller}) = 0.8$
      $Pr_w(\textbf{anti-nuclear-protests} \mid \textbf{feminist-teller}) = 0.7$

---

might well be good-enough explanation for the facts in the description, while by no means being *the only* good-enough explanation. The alternative set under consideration will matter greatly: any speaker who is restricting attention to the explicitly given alternatives ("bank teller" and "bank teller who is active in the feminist movement") should be happy to consider the conjunctive alternative as the only good-enough explanation. But a speaker who also considers "active in the feminist movement" as an alternative, plausibly generated from the conjunctive alternative via deletion (Katzir, 2007), should in fact consider the *must* statement in (29b) as plain false. We suspect that the version with *ought* (or *should*) will be far more acceptable to these speakers because the semantics of *ought* as a weak necessity modal does not have the same exhaustification component as *must* in our approach, in ways we outline in section 3.1.

[20]We thank a reviewer for pressing us on this matter.

(32)   The explanatory value of Linda being a bank teller

$$\mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{teller}] = 0.3 + 0.2 = 0.5$$

(33)   The explanatory value of Linda being a feminist bank teller

$$\mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{feminist-teller}] = 0.8 + 0.7 = 1.5$$

Given that the explanatory value of the hypothesis **feminist-teller** is (significantly) greater than the explanatory value of the hypothesis **teller**, one is led to conclude that the former hypothesis is the only good explanation of the evidence among the salient hypotheses.

(34) $[\![$ must **feminist-teller** $]\!]^w$
$$= (\mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{feminist-teller}] > \theta) \wedge (\mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{teller}] \leq \theta)$$

If (29b) constitutes a modal conjunction fallacy as we strongly suspect, our theory explains it fully and immediately, while building on tools from formal epistemology that have been applied very successfully to the psychology of reasoning.

The conjunction fallacy plays only a supporting role in our thesis in this article. First, it demonstrates that confirmation-theoretic mechanisms such as our proposal for the semantics of necessity epistemics are part of higher cognition. If we see evidence of confirmation theory in deliberate reasoning, it should not strike us as too alien to find it in the meaning of some modal expressions in natural language. Second, our theory of necessity epistemics immediately predicts the existence of modal versions of the conjunction fallacy, demonstrating its generative power.[21,22]

---

[21] Thanks to comments by an anonymous reviewer, we realized that our theory of *must* suggests a more ambitious possible account of the *original* conjunction fallacy as discovered by Tversky and Kahneman (1983). Various scholars in philosophy of language and formal pragmatics have proposed assertion operators that would apply in a systematic manner to declaratives meant to impart information. The pragmatic version of this move sees it not as a covert operator in logical form, but as an inference, for example: if the speaker uttered assertion φ, then the speaker *believes φ* (Grice, 1975; Sauerland, 2004; Stalnaker, 1978). Other, non-pragmatic approaches postulate explicit operators in logical form (Meyer, 2013). All such proposals we are familiar with postulate box-type modal operators, that is universal quantifiers over possible worlds or situations, analogous to strong necessity modals like *must*. If this class of proposals is on the right track, it is interesting to consider whether the semantics for *must* we give in this article might be a reasonable contender for such an assertion operator. In the event that it is, the original conjunction error could be explained as a result of such a silent assertion operator's having the kind of semantics we propose here for the overt English modal *must* and its overt counterparts in other languages. Even more ambitiously, one would ask whether other instances of unexpected confirmation-theoretic inference-making behavior in the psychology of reasoning find their root in such a silent assertion operator. We cannot present a careful consideration of this theoretical possibility in this article, and must leave it for later research.

[22] An anonymous reviewer points out that our theory will have trouble explaining the so-called *A-B paradigm*, which contrasts with the Linda problem in that it does not introduce a context establishing a psychologically salient connection with one of the hypotheses. Given the following task, Tversky and Kahneman report that 58% of the participants considered the conjunction ($h_1 \wedge h_2$) more probable than one of its conjuncts (h_1).

(i)   A health survey was conducted in a representative sample of adult males in British Columbia of all ages and occupations. Mr. F. was included in the sample. He was selected by chance from the list of participants.

Which of the following statements is more probable?
a.   Mr. F. has had one or more heart attacks. (h_1)
b.   Mr. F. has had one or more heart attacks and he is over 55 years old. ($h_1 \wedge h_2$)

We acknowledge that our theory, which is more or less in line with Crupi et al.'s (2008) analysis based on inductive confirmation, is not immediately applicable to the A-B paradigm due to the presupposition of the existence of relevant pieces of information/evidence. We have two remarks on the A-B paradigm. First, people make significantly more mistakes

3.3 Lawyers and engineers (Kahneman & Tversky, 1973)

Kahneman and Tversky (1973) argue that human reasoners neglect prior probabilities when solving ostensibly probabilistic problems, relying instead on judgments of typicality. In the "lawyers and engineers" experiment, subjects were asked to provide the probability of Jack being an engineer based on the description in (35).

(35)   A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written. Below is the thumbnail description of Jack, one of the interviewees:

   Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.

Kahneman and Tversky tested two conditions between participants: in one, Jack's description was drawn randomly from a sample of 30 engineers and 70 lawyers, as in (35) above. In the other condition the prior probabilities were reversed, and the sample consisted instead of 30 lawyers and 70 engineers. They found that participants' judgments were unaffected by these prior probabilities: participants in the 30–70 condition gave the same response to the question about the probability that Jack is an engineer as participants in the 70–30 condition. This suggests that indeed they were not resorting to the normative standard provided by Bayes' theorem to decide on their response.[23]

   In our introspection, it seems possible to replicate the issue with modalized expressions. Given the same description of Jack, upon being asked to guess whether Jack is a lawyer or an engineer, it is reasonable to utter the following:

(36)   Jack must be an engineer.

Our prediction is that naive human participants would prefer (36) to an alternative "Jack must be a lawyer". This is just as surprising as the reported result in the original experiment: to assent to (36) in such an experiment is to display a semantics for *must* that is not as sensitive to prior and posterior probabilities as extant probabilistic semantics for *must* would predict.

---

in the Linda case (85%) than in the aforementioned heart attack case (58%). This casts doubt on the view that a single factor is solely responsible for the conjunction fallacy in all its variants. Second, there is an alternative inductive confirmation-based explanation due to Tentori et al. (2013), which does *not* compare the degrees to which given evidence confirms two salient hypotheses, but rather, directly measures the degree to which one conjunct (h_1) inductively confirms the other (h_2). Tentori et al. experimentally verify that this measure is a good predictor of the A-B paradigm. But such innovation comes at a cost; it loses the original appeal of Crupi et al.'s (2008) theory, namely that people compare the inductive confirmatory values of competing hypotheses.

[23]The two between-subjects conditions are of the essence, as the mere fact that priors have little to no effect in one condition is not enough to argue that elements of Bayes' theorem are being ignored. This is because an extreme likelihood term (probability of the description of Jack assuming he is an engineer) will have the effect of diluting the role of priors determining posteriors, following Bayes' theorem. But such an extreme likelihood term should then be visible in the other condition, where priors were flipped: Bayes' theorem would lead us to expect an even higher posterior probability for engineer, as long as responses were not at ceiling, which indeed they were not. Instead, Kahneman and Tversky (1973) found that participants in the two conditions gave indistinguishable responses, providing a compelling case that in the lawyers-and-engineers task as administered in this experiment, participants are indeed ignoring prior probabilities.

We argue that the explanatory value of **engineer** with respect to the provided description is greater than the explanatory value of **lawyer**, and that the prior probabilities of the two hypotheses have little to no direct effect on such a calculation of explanatory adequacy.[24]

To illustrate the mechanics of our account, we will consider in detail the two most relevant pieces of information about Jack, namely that he shows no interest in political and social issues and enjoys solving mathematical puzzles. Below is what we deem to be reasonable probability assignments regarding the two crucial pieces of evidence:[25],[26]

(37) $Pr_w(\textbf{not-political-social} \mid \textbf{engineer}) = 0.78$
         $Pr_w(\textbf{enjoys-mathematical-puzzles} \mid \textbf{engineer}) = 0.55$

(38) $Pr_w(\textbf{not-political-social} \mid \textbf{lawyer}) = 0.35$
         $Pr_w(\textbf{enjoys-mathematical-puzzles} \mid \textbf{lawyer}) = 0.28$

The probability of Jack showing no interest in political and social issues given that he is an engineer is 0.78, and the probability of him enjoying mathematical puzzles given the same hypothesis is 0.55. By contrast, the probabilities of Jack showing no interest in political and social issues and him enjoying mathematical puzzles given that he is a lawyer are 0.35 and 0.28, respectively.

(39)   The explanatory value of Jack being an engineer

$$\mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{engineer}] = 0.78 + 0.55 = 1.33$$

(40)   The explanatory value of Jack being a lawyer

$$\mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{lawyer}] = 0.35 + 0.28 = 0.63$$

---

[24]The illustration we give here uses our likelihoodist semantics for *must*, which altogether ignores prior probabilities. Other, more sophisticated Bayesian measures of confirmation show non-zero degrees of sensitivity to prior probabilities, and might make for a more complete account.

[25]The probabilities were taken from a norming study on the lawyers and engineers scenario (Guerrini, Sablé-Meyer, & Mascarenhas, 2022).

[26]Tversky and Kahneman (1983) and Tentori et al. (2013) reject likelihoods as the relevant measure due to the following Wimbledon scenario:

(i)   Suppose that Bjorn Borg reaches the Wimbledon finals in 1981.
        a.   Borg will lose the first set.
        b.   Borg will lose the first set but win the match.

Tversky and Kahneman report that people judge (i-a) more likely than (i-b), but this cannot be explained in terms of likelihoods. They argue that "it makes no sense to assess the conditional probability that Borg will reach the finals given the outcome of the final match". Tentori et al. make a similar point: "the inverse probability analysis must imply the utterly implausible judgmental strategy of focusing on the probability of Borg's Wimbledon record, which is in fact an established datum from the past, as conditional on future events concerning the outcome of the final match".

We think that their dismissal was too hasty. As weird as it may seem, the relevant likelihood is mathematically well-defined. There is no problem treating the alleged future tense *will* as a modal, and in this case, the relevant likelihood measure merely conditions on a modal rather than a future event. Moreover, in economics, conditioning on future events is a widely used methodology. For instance, in a time series analysis, it is common to calculate $Pr(X(1) > 10 \mid X(2) > 30)$ where $X(t)$ denotes a stock price at time t and the current time is 0. We thank Janek Guerrini for discussion of this argument.

Given the above probability assignments, 'Jack must be an engineer' is true if and only if 'the hypothesis that Jack is an engineer is the only good-enough explanation of the given evidence among the candidate hypotheses'.

(41) $[\![\text{ must } \mathbf{engineer}]\!]^w = (\mathbb{E}_w[\mu_{\text{EVAL}} \mid \mathbf{engineer}] > \theta) \wedge (\mathbb{E}_w[\mu_{\text{EVAL}} \mid \mathbf{lawyer}] \leq \theta)$

4 Natural language evidence: conditional evaluatives

In this section, we compositionally derive our proposed semantics from Korean conditional evaluatives (repeated below as (42)), which have a transparent morphosyntax.

(42)   John-un cip-ey   iss-∅-**eya**   **toy**-n-ta.
         John-TOP home-DAT COP-PRES-**only.if** EVAL-PRES-DECL
             '(Lit.) Only if John is home, it suffices.'
         'John must be home.'

We conjecture that the above conditional evaluative construction is the transparent version of the English necessity modal *must*. Despite the fact that modal necessity is expressed via an auxiliary in English but via a full-fledged conditional construction in Korean, we conjecture that their meanings more or less converge for the following reason: People's understanding of obligation/permission/utility (deontic) or probability (epistemic) is rather consistent regardless of their mother tongue; otherwise we would expect abundant communication failures between native speakers of different languages in a modal talk.[27] And since modal expressions are precisely the means to convey such concepts, it is reasonable to assume that English and Korean modal expressions convey similar meanings.[28]

For a compositional analysis, we will break down the conditional evaluative into three subcomponents: (i) the evaluative predicate, (ii), the conditional, and (iii) the exhaustifier. We first show that composing the first two subcomponents yields an expected utility measure for deontics and a likelihood-based confirmation measure for epistemics. The exhaustifier is responsible for comparing the relevant measures.

4.1 Deriving relevant measures from conditional semantics

We assume that the evaluative predicate *toy* 'EVAL' is a measure function with the semantics already presented in (7), repeated below as (43).[29]

---

[27]More bluntly put: Suppose that you—a native speaker of English—are advising an international student. When you give directions, do you expect the student to accidentally disobey the order because they have a different understanding of obligation as a non-native speaker of English?

[28]We do not intend to claim that all modal expressions across languages convey exactly the same meaning. In fact, Deal (2011) argues that Nez Perce does not lexically distinguish modal necessity from modal possibility, and this is evidence that we cannot always find a one-to-one correspondence of modal expressions in any given pair of languages.

[29]We gloss Korean *toy* as 'EVAL' to emphasize its bleached status. In other contexts, the morpheme seems to convey the meaning of 'suffice', as exemplified below:

(i)   kule-myen toy-n-ta.
        do.so-if   EVAL-PRES-DECL
        'It suffices to do so.'

(43) *toy* 'EVAL' as a measure function over worlds

$$[\![\text{EVAL}]\!]^w = \mu_{\text{EVAL}} = \lambda w'. \mid \{r \in R \mid r(w') = 1\} \mid,$$

where R is the set of relevant propositions

As for the semantics of conditionals, we assume that conditionals denote the degree of support for the consequent, given the antecedent. Technically, the value of 'if φ then γ' is the expected value of γ given φ.[30,31]

(44) $$[\![\text{if } \varphi, \text{ then } \gamma]\!]^w = \mathbb{E}_w[\gamma \mid \varphi] = \sum_{w_j \in \varphi} \gamma(w_j) \times Pr_w(\{w_j\} \mid \varphi)$$

Note that when the value of the consequent is either 0 (false) or 1 (true), the expected value reduces to the *probability* of the consequent given the antecedent; the probability-weighted average of γ given φ is by definition the conditional probability of γ given φ. This proves that conditional probability is a special case of expected value, and it follows that the posited semantics is in accordance with Adams (1965), Douven (2008), and Pearl's (2000, 2013) analyses of conditionals (see also Crupi & Iacona, 2022; Gibbard, 1980; Jackson, 1979; Jeffrey & Edgington, 1991; S. Kaufmann, 2005; Lewis, 1976, for relevant work in linguistics and philosophy). However, we depart from previous work in that we do not restrict the type of the consequent of conditionals to propositions. This is particularly important for our analysis because the consequent of Korean conditional evaluatives is not a proposition but rather a measure function.

To derive the proposed measure, we simply have to replace the consequent γ of the conditional in (44) with the evaluative predicate *toy* 'EVAL'. Note that this yields exactly what we proposed in (9) and (12). We take this as natural language evidence that such a measure is employed by at least some modals.

(45) $$[\![\text{if } \varphi, \text{ then EVAL}]\!]^w = \mathbb{E}_w[\mu_{\text{EVAL}} \mid \varphi] = \sum_{w_j \in \varphi} \mu_{\text{EVAL}}(w_j) \times Pr_w(\{w_j\} \mid \varphi)$$

---

[30] We want to make it clear that we are not claiming that (44) is precisely what conditionals denote. It suffices for our purposes to adopt the simplest formulation among extant expected value-based theories of conditionals. We leave it open as to whether the skeleton of our theory can be made compatible with a more nuanced semantics such as Douven (2008) or Crupi and Iacona (2022).

[31] An anonymous reviewer asks whether it is possible to give an expected value-based analysis of subjunctive conditionals while capturing the seemingly close connections between indicatives and subjunctives. Pearl (2000) develops a probabilistic analysis of subjunctive conditionals which builds on his theory of causation. In this view, subjunctive conditionals are interpreted with respect to a network of causally relevant variables. Abstracting away from the details, a counterfactual assumption φ severs the causal connection between the variable related to φ and its causes (i.e., *intervenes* on φ). Given this modified network, a subjunctive conditional '$\varphi \;\square\!\!\rightarrow\; \psi$' denotes the expected value of ψ conditionalized on φ and its causally relevant, true propositions. Pearl notes that the crucial difference between subjunctives and indicatives is whether the interpretive process involves intervening on the antecedent or merely observing the truth of the antecedent. While Pearl's theory departs from Lewis's (1973) similarity-based semantics in that causation is taken as a primitive, there have been attempts in linguistics and philosophy to incorporate Pearl's insights: S. Kaufmann (2005) offers an analysis of subjunctives in terms of expected value and causal structure. Moreover, Schulz (2011), S. Kaufmann (2013), Ciardelli, Zhang, and Champollion (2018), and Santorio (2019) modifies premise semantics (Kratzer, 1979) in such a way that it is sensitive to a causal structure.

Note that the conditional denotes a degree rather than a proposition. Following Lassiter (2017), we suggest that a degree representation can be mapped to a bivalent one by invoking the thresholding operation.[32]

(46)   Thresholding operator (Lassiter, 2017)

   $\Theta = \lambda d.d > \theta,$ where θ is a contextually determined threshold

Feeding the denotation of the conditional to Θ yields the semantics in (47) informally read as follows: the conditional is true if and only if the measured value of φ is greater than the contextually determined threshold θ.

(47) $\Theta(\llbracket$ if φ, then EVAL $\rrbracket^w) = \mathbb{E}_w[\mu_{\text{EVAL}} \mid \varphi] > \theta$

We are only half through composing the semantics of the conditional evaluative construction, as we have not considered the exhaustification component of *-(e)ya* 'only if' yet. In what follows, we claim that the exhaustification component indirectly compares the measured value of φ to the measured values of its contextually salient alternatives.

4.2 Exhaustification

We simply assume that the exhaustification component of *-(e)ya* 'only if' takes a proposition φ and negates each of its alternatives, along with conveying that φ is true.[33] This is exactly what we proposed for the analysis of modal necessity in (6).

(48)   The compositional semantics of Korean conditional evaluatives

$$\llbracket \text{ only}_{\text{-(e)ya}} \rrbracket^w (\Theta(\llbracket \text{ if } \varphi, \text{ then EVAL } \rrbracket^w))$$
$$= (\mathbb{E}_w[\mu_{\text{EVAL}} \mid \varphi] > \theta) \wedge \forall \psi \in Alt(\varphi) : \neg(\mathbb{E}_w[\mu_{\text{EVAL}} \mid \psi] > \theta)$$
$$= (\mathbb{E}_w[\mu_{\text{EVAL}} \mid \varphi] > \theta) \wedge \forall \psi \in Alt(\varphi) : (\mathbb{E}_w[\mu_{\text{EVAL}} \mid \psi] \leq \theta)$$

Hence we have independent evidence from natural language that a decision theoretic notion of expected utility and Bayesian confirmation theoretic measures are relevant to the interpretation of linguistic modality.

5 Prior probabilities and the problem of success

One of the key features of our theory in its current form is that modal interpretation ignores the prior probabilities of the prejacent and its alternatives. While this insensitivity to priors matches intuitions at multiple empirical junctures, and allows our theory to address puzzles of *failure* of reasoning (i.e., why do people make fallacious

---

[32]The thresholding operation is reminiscent of the *pos* morpheme of Kennedy and McNally (2005) and Kennedy (2007). We make a distinction between Θ and *pos* only because Θ is a function from degrees to truth values whereas *pos* is of a higher order type due to compositional issues. Apart from the type-related concern, no part of our analysis hinges on making such a distinction.

[33]We remain agnostic on whether conveying that φ is true is a presupposition or an assertion, as there is no evidence that the exhaustification component of *-(e)ya* 'only if' behaves exactly like English *only*. Besides, we would like to focus on the formulation of modality, due to reasons of space.

inferences?), it naturally raises a question as to how the theory can explain the puzzle of *success* (i.e., how can people make classically sound inferences despite all?).

The puzzles of failure and success are the two sides of the human reasoning coin, and it is unusual for a theoretical approach to answer both questions in comparable terms. In particular, linguists and philosophers have traditionally focused on the puzzle of success, while psychologists mostly paid attention to the puzzle of failure. What we presented in earlier sections is a linguistic theory of the meaning of *must* that predicts what might look like failures of reasoning, based on a novel modal semantics. For the remainder of this section, we give tentative directions as to how the puzzle of success can be considered within the spirit of our theory.

Let us first note that the lack of an extensive explanation of the puzzle of success does not immediately provide sufficient grounds to reject our theory. Just as much as our theory suffers from the puzzle of success, alternative theories that build on priors have trouble handling the puzzle of failure and need to stipulate that people often ignore priors for extrinsic and often mysterious reasons. While reasoning experiments do not seem to favor a particular theory, we have good evidence that at least some modals are interpreted in the way we proposed: analyzing the Korean modal data in a what-you-see-is-what-you-get manner yields the expected value-based semantics.

However, priors clearly *can* factor into modal reasoning. Consider the example in (49). Upon hearing that John did not come to work, one could reasonably conjecture that he must have caught a cold. By contrast it is infelicitous to say that he must be dead, despite the fact that his being dead would fully predict and explain the relevant fact that he is absent.[34]

(49)   John did not come to work today.
  a.   He must have caught a cold.
  b.   #He must be dead.

Different measures of hypothesis testing make different predictions regarding this example, but let us focus on the ones relevant to our theory. In terms of likelihoods, the hypothesis that John is dead is the best explanation of his being absent since $Pr(\textbf{absent} \mid \textbf{dead}) = 1$. This hypothesis remains attractive even in view of the likelihood ratio measure, as $Pr(\textbf{absent} \mid \textbf{dead}) \gg Pr(\textbf{absent} \mid \neg\textbf{dead})$. Given its strong preference for the hypothesis that John is dead, our core theory as it stands incorrectly predicts that (49a) is false whereas (49b) is true. Note that the prediction remains unaltered even if one entertains a different alternative to **dead** such as 'John caught a cold', as $Pr(\textbf{absent} \mid \textbf{dead}) \gg Pr(\textbf{absent} \mid \textbf{cold})$.

(50)   Likelihood-based comparison

$$Pr(\textbf{absent} \mid \textbf{dead}) = 1 \gg Pr(\textbf{absent} \mid \neg\textbf{dead})$$

---

[34]We thank Benjamin Spector for pointing out to us this prediction of our theory.

27

(51)   Likelihood ratio-based comparison

$$L(\textbf{dead}, \textbf{absent}) = \log\left(\frac{Pr(\textbf{absent} \mid \textbf{dead})}{Pr(\textbf{absent} \mid \neg\textbf{dead})}\right)$$

One could opt for other Bayesian measures of confirmation that are sensitive to priors such as the D measure introduced in Section 1.2. Recall that D is the difference between the posterior probability and the prior. While still making the right predictions for the conjunction fallacy, the D measure penalizes hypotheses with extremely low priors and posteriors. Let us illustrate with plausible probability assignments:

(52)   Plausible probability assignments

$$Pr(\textbf{dead} \mid \textbf{absent}) = 0.001, \quad Pr(\textbf{dead}) = 0.0001$$
$$Pr(\textbf{cold} \mid \textbf{absent}) = 0.7, \quad\;\; Pr(\textbf{cold}) = 0.1$$

(53)   Corresponding D measures

$$D(\textbf{dead}, \textbf{absent}) = Pr(\textbf{dead} \mid \textbf{absent}) - Pr(\textbf{dead})$$
$$= 0.001 - 0.0001 = 0.0009$$
$$D(\textbf{cold}, \textbf{absent}) = Pr(\textbf{cold} \mid \textbf{absent}) - Pr(\textbf{cold})$$
$$= 0.7 - 0.1 = 0.6$$

According to the above probability assignments, $D(\textbf{cold}, \textbf{absent})$ is significantly greater than $D(\textbf{dead}, \textbf{absent})$, primarily due to the fact that the prior and posterior of **dead** are extremely low. Consequently, the difference between the prior and the posterior is minute.

Despite the appeal, there is one serious drawback to employing such a measure: we would lose the established parallelism between deontic and epistemic modals. Recall that expected utilities and likelihoods are derived exactly in the same manner and this was part of the motivation for our analysis of epistemic modality. But we see no simple way of similarly deriving expected utilities and the D confirmation measure from one and the same core definition. Since this connection remains at the heart of our theory, we must seek alternative routes to account for the sensitivity to priors.

We suspect that the best way to capture the contrast in (49) is to require that the prior probability of the modal prejacent is reasonably high, although it need not be higher than the prior probabilities of its alternatives. This requirement would be entirely independent of the particular modal domain, in keeping with our goal to give a core semantics for *must*. That is, a sufficiently high prior probability would be a requirement for epistemic, deontic, and other modalities. Such a requirement can be viewed intuitively as a plausibility requirement: whether the statement 'must φ' is epistemic or deontic or teleological, the proposition φ had better be plausible or feasible.

In the epistemic domain, this requirement makes (49a) a reasonable thing to say because a cold is quite common a condition and accordingly has a relatively high prior. By contrast, (49b) is false or infelicitous because **dead** is extremely unlikely in a normal context. Accordingly, the sentence improves if John's country of residence is in a war situation and his neighborhood is bombarded on a regular basis, or if John is very old.

This view makes the following prediction regarding the lawyers and engineers scenario: if the group of interviewees consists of 99 lawyers and 1 engineer, one would be reluctant to accept 'Jack must be an engineer' for the same reason that 'John must be dead' sounds odd in a normal context. In fact, there are reports in the psychology literature that priors are more diagnostic when they have extreme values (Koehler, 1996; Ofir, 1988; Wells & Harvey, 1977).

In the deontic and teleological domains, the requirement translates naturally as a requirement of plausibility/feasibility.[35] Thus, (54a) and (54b) would be infelicitous or plain false (more on which below), showing that the requirement extends to weak necessity modals. Similarly for (55).

(54)   a.   # One must under no circumstance ever make a grammatical mistake.
            b.   # One should under no circumstance ever make a grammatical mistake.

(55)   a.   #In order to get to Bushwick, you have to take a helicopter.
            b.   #In order to get to Bushwick, you should take a helicopter.

What is the status of this plausibility requirement? We have somewhat conflicting judgments. The sentences with strong necessity modals strike us as plain false: in order to get to Bushwick, it is simply not the case that you have to take a helicopter, for there are multiple alternative ways of accomplishing your goal, irrespective of the impracticality of the helicopter alternative. This suggests that the requirement should be seen as an entailment affecting the truth conditions of the sentence. Accordingly, the negated sentences in (56) seem felicitous and true.

(56)   a.   It's not the case that one must under no circumstance ever make a grammatical mistake.
            b.   In order to get to Bushwick, you don't have to take a helicopter.

Yet, some not-at-issue projective content is happy with negation: "The king of France was not in attendance at the party last night" isn't too hard to read as plain true. Additionally, it is hard to disentangle propositional negation, which is what we intend in the sentences in (56), from its meta-linguistic uses, at least when targeting presuppositions. We thus find the data from (56) at best suggestive of an at-issue, non-projecting content analysis of the plausibility requirement.

To our ears, the interrogative versions of these sentences can be addressed in dialog with negation, but the hey-wait-a-minute construction strikes us as entirely appropriate:

---

[35]Somewhat relatedly, Lassiter's (2011) analysis of deontic *must* compares the expected utilities of only those alternatives that are sufficiently plausible. Sloman (1970) also notes that *ought* statements only pay attention to practical possibilities, i.e., possibilities that can be brought about.

(57) a. Q: Must one under no circumstance ever make a grammatical mistake?
          A: No, you can make mistakes occasionally.
          A: Wait a minute, that's practically impossible anyway!
      b. Q: In order to get to Bushwick, do I have to take a helicopter?
          A: No, there are other means of getting there.
          A: Wait a minute, that's not a plausible option anyway!

Similarly for the epistemic domain:

(58) John didn't show up today,
      a. It's not the case that he must be dead.
      b. Q: Must he be dead, do you think?
          A: No, he probably just has a cold.
          A: Wait a minute, that's an implausible possibility!

Let us take stock of this section. We argued that our proposal makes sense of seeming rationality violations with *must*: where it looks like humans are erroneously ignoring prior probabilities, we say that they are doing so rationally, because modal operators in the epistemic domain are not about maximizing posterior probability, but rather explanatory power. However, our proposal gets into trouble for predicting *no* effect of prior probabilities whatsoever across the board in the epistemic domain. This view is clearly too radical, and must be tempered somehow. One large domain of possibilities is to use Bayesian confirmation measures (i.e., not simply the likelihood of the prejacent), for in many of these measures the prior probabilities play a role, as we illustrated with the *D* measure, which subtracts the prior probability from the posterior. This avenue is extremely promising for the epistemic case, but it seems it would defeat one of the central goals of our work in this article, namely to give one and the same fundamental semantics for modals, irrespective of modal domain.[36] So, we proposed instead that prior probabilities play a role in the form of a plausibility requirement: the prejacent must have a prior probability above some contextual standard for plausibility. We showed how this proposal handles the problematic epistemic cases and makes reasonable predictions on the expected-utility side. We could not determine the exact nature of this requirement, in particular whether it is standard truth-conditional content or projective content. On the one hand, family-of-sentences tests suggest that this content does not project. On the other, some kinds of projective content, for example definite descriptions, are easy enough to "trap" inside truth conditions under negation and other operators, and we've shown that it is entirely appropriate to react to the modal sentences in question by targeting the plausibility requirement as one would target say a factive presupposition.

---

[36]We are not saying that this strategy is in principle incompatible with our goal of having a single lexical entry for each modal. We see no immediate reason to suspect so. But it is clear that the simple, intuitive lexical entry we propose in this article for *must* cannot be straightforwardly adapted to work with the *D* measure without bringing serious issues on the deontic (utility) front. But of course this is not to say that it is impossible to give such an entry, or that there aren't other Bayesian confirmation measures that would solve the issues on the epistemic side without destroying our results on the deontic side.

6 On the interpretation of possibility modals

Thus far, we developed a semantics for so-called necessity modals. A natural question to ask is how possibility modals such as *might* or *may* relate to necessity ones: as an anonymous reviewer points out, we want to systematically rule out statements such as "It must be raining, but of course it might not be". The Kratzerian account and modal logic capture this by assuming that necessity and possibility modals are duals, e.g., 'might φ' is equivalent to '$\neg$must $\neg\varphi$'. Assuming duality in our theory yields the following semantics:

(59) Epistemic possibility

$$[\![\text{ might } \varphi \ ]\!]^w = [\![\ \neg \text{ must } \neg\varphi \ ]\!]^w$$
$$= \neg((\mathbb{E}_w[\mu_{\text{EVAL}} \mid \neg\varphi] > \theta) \wedge \forall \psi \in Alt(\neg\varphi) : (\mathbb{E}_w[\mu_{\text{EVAL}} \mid \psi] \leq \theta))$$
$$= (\mathbb{E}_w[\mu_{\text{EVAL}} \mid \neg\varphi] \leq \theta) \vee \exists \psi \in Alt(\neg\varphi) : (\mathbb{E}_w[\mu_{\text{EVAL}} \mid \psi] > \theta)$$

The formula reads as "might φ is true if and only if the explanatory value of $\neg\varphi$ is *not* sufficiently high, or there exists an alternative to $\neg\varphi$ such that its explanatory value is sufficiently high". We find this a reasonable proposal for the meaning of *might*. Consider the first disjunct: if $\neg\varphi$ is not sufficiently explanatory then we do not have sufficient grounds to reject φ, hence it is possible that φ is true. Regarding the second disjunct, we first observe that, while the literature has presented arguments in favor of the idea that *must* is sensitive to alternatives, we are unaware of such arguments in favor of alternative sensitivity of *might*. Adapting a scenario from Dretske (1972), imagine that Kim will only inherit the considerable fortune their parents left them if they get married. They can marry anyone they like, the condition is simply that Kim be married in order to inherit. Suppose Kim is planning on marrying Pat, and consider the *must* sentences in (60), where SMALL CAPS indicate focus.

(60) a. Kim must marry Pat in order to inherit.
b. Kim must marry PAT in order to inherit.

There is a clear contrast: sentence (60a) is either true or true enough, while sentence (60b) is plain false. Any reasonable alternative-sensitive approach to *must* accounts for this. On our proposal, (60a) with neutral focus plausibly contrasts the prejacent with its negation as an alternative, yielding truth, while focus in (60b) strongly suggests a question under discussion concerning other individuals Kim might marry, and is accordingly predicted to be plain false, since marrying Pat is by no means the only good-enough course of action given the stated goals, when the alternatives concern other individuals Kim might marry. Crucially, no such contrast is to be found with analogous *might* sentences:

(61) a. Kim might marry Pat in order to inherit.
b. Kim might marry PAT in order to inherit.

To our ears, (61b) sounds a little odd, since one can't quite make out what justifies the focus on PAT. But there is no truth-conditional contrast between the two sentences. We

conclude from these facts that there is no evidence in favor of alternative sensitivity for *might*, at least not of the same kind as the alternative sensitivity of *must*. Vitally, this is not to say that the *semantics* and the *truth conditions* of *might* sentences have nothing in them that formally corresponds to an alternative set. It only means that the alternatives of *might*, if there are any, cannot be manipulated by context, or can be manipulated but never make a difference for truth conditions. With this in mind, we propose that the expression $Alt(\varphi)$ that occurs in (59) is in fact non-manipulable, and is fixed as the polar alternative to φ. The second disjunct of our entry in (59) then says that a sentence of the shape 'might φ', analyzed as '¬ must ¬φ', will be true if the alternative to the prejacent ¬φ, namely φ, has a sufficiently high explanatory value, which indeed is a good reason to accept 'might φ'.

It is interesting to note that the alternative analysis we considered in Section 2, which directly makes reference to the L confirmation measure (cf. (14)) offers a perhaps even more intuitive interpretation of *might*:

(62)   Epistemic possibility: an alternative analysis

$$
\begin{aligned}
[\![\text{ might}_{\text{LOG,ALT}} \, \varphi \,]\!]^{w} &= \neg(L(\neg\varphi, e) > \theta) \\
&= L(\neg\varphi, e) \leq \theta \\
&= log\left(\frac{Pr_w(e \mid \neg\varphi)}{Pr_w(e \mid \varphi)}\right) \leq \theta \\
&= log(Pr_w(e \mid \neg\varphi)) - log(Pr_w(e \mid \varphi)) \leq \theta \\
&= log(Pr_w(e \mid \varphi)) - log(Pr_w(e \mid \neg\varphi)) \geq -\theta \\
&= log\left(\frac{Pr_w(e \mid \varphi)}{Pr_w(e \mid \neg\varphi)}\right) \geq -\theta \\
&= L(\varphi, e) \geq -\theta
\end{aligned}
$$

The above formula states that 'might φ' is true if and only if the L confirmation measure of φ is greater than the contextually determined threshold $-\theta$. Recall that positive values indicate positive confirmation, negative values signify negative confirmation, and deviation from 0 by θ conveys significance. So intuitively, the formula conveys that 'might φ' is true if and only if φ is not significantly disconfirmed. Thus in our alternative analysis, *must* and *might* concern significant confirmation and lack of significant disconfirmation, respectively. This perspective has the cost of oversimplifying the semantics: since the set of relevant alternatives exclusively consists of the prejacent and its negation even in the *must* case, this view effectively renders the semantics insensitive to more interesting alternative sets. While there might be reasons to endorse this insensitivity to alternatives in the epistemic domain (e.g., Yalcin's (2005) argument concerning Kyburg's (1961) lottery scenario), it would be largely inadequate in the deontic domain. We leave further development for future work.

7 Further implication: the weakness of epistemic necessity

The theory of strong necessity modals we offered here generates a rather weak interpretation of *must* in the epistemic domain, in that a proposition φ needn't have a high probability for 'must φ' to be true. Rather, what matters is the explanatory value of φ with respect to a salient body of evidence. How does our account deal with other arguments for a weak semantics for *must*?

In a now classic article on necessity modals, von Fintel and Gillies (2010) establish an important puzzle for strong semantics for *must*, which we've nodded to at multiple points in this article. They point out that there is a contrast between (63) and (64), and submit that this is because, in (63), Billy directly obtained the information that it is raining, while in (64) this information was indirectly acquired.

(63) *Billy is looking out the window at the pouring rain.*

   #It must be raining.

(64) *Billy sees someone enter the building holding a wet umbrella, but she herself cannot see outside.*

   It must be raining.

In this article, we proposed that epistemic 'must φ' asserts that φ is the only good-enough explanation for a contextually determined, salient body of evidence. In (64), the context makes it clear that the evidence to be explained is the fact that someone just came in with a wet umbrella. An event of **rain** would be an excellent explanation for that fact, and our account predicts this: presumably, conditional on **rain**, the probability of a wet umbrella for someone who was just outside is extremely high, and no alternative pops to one's mind in this bare-bones context. The case of (63) is more interesting, for there the salient evidence to explain is rain itself. Formally, the probability of **rain** conditional on **rain** is, of course, as high as any probability can get. As discussed so far then, our account technically predicts that (63) should be a true and felicitous sentence. However, the analytical intuition behind our account, as we've explained in detail above, is that epistemic *must* is about *explanatory power*. And a proposition φ is no explanation or argument for φ itself, this is a clear instance of question begging.

We propose to rule out cases of checking probabilities of the shape $Pr(\varphi \mid \varphi)$ for pragmatic reasons, essentially a probabilistic version of the pragmatic principles that generate infelicity for tautological sentences in a bivalent semantics. For notice that our predicted truth conditions for (63) are "the probability of **rain** conditional on **rain** is above the threshold θ, and none of the probabilities of alternatives to **rain** are above θ". The second clause of these truth conditions isn't exactly trivial,[37] but the first clause requires that we consider a probability of the shape $Pr(\varphi \mid \varphi)$, which we

---

[37]In principle, it might be possible to weave a context where **rain** is the evidence to be explained, but there are two alternative explanations, **rain** versus ψ, where the probability of rain conditional on ψ is also above the threshold θ. Since thresholds are hard or impossible to manipulate with any precision, at least with the tools of introspection, we cannot decide here whether this configuration can be induced while preserving coherence. See also the discussion at the end of Section 3.1 on how easily thresholds must be allowed to shift in order to account for the miners puzzle in exhaustive semantics for *must* like ours.

would expect to trigger infelicity. To be clear, our view here is not that "it must be raining", in the context at hand, is a trivial, tautological sentence. Rather, the sentence is deviant because it crucially involves the at-issue assessment of a trivial probability of the shape $Pr(\varphi \mid \varphi)$. Zooming out, this sensible constraint will rule out any *must* statement where the prejacent *entails* the evidence to be explained. This is as intended, and meant to block question-begging (non-)explanations.

Above and beyond this natural pragmatic requirement for non-trivial explanations, our proposal captures the idea that any epistemic *must* sentence with a known prejacent should be infelicitous (Giannakidou & Mari, 2016; Goodhue, 2017), for it considers alternatives to the prejacent as possible antecedents to conditionals, in a manner we elucidate presently.

Goodhue notes that from the perspective of a skeptical epistemologist, 'it must be raining' can be felicitous even when she observes the pouring rain, as in (65).

(65)   *A professional epistemologist, while on vacation in Seattle, looks out the window at the pouring rain. She says:*
        It must be raining.

Goodhue proposes that this context dependency of the felicity condition can be accounted for if '*must* φ' requires that φ is not known and Lewis's (1996) context dependent theory of knowledge is adopted:

(66)   Lewis (1996) on knowledge
        The speaker knows that $\varphi \leftrightarrow$ The speaker's evidence eliminates every possibility in which $\neg\varphi$—except for those possibilities that we are properly ignoring.

In this view, the professional epistemologist does not deduce that it is raining from observing the pouring rain outside the window, because she considers far-fetched possibilities where it does not rain despite her observing the rain (e.g., she has a delusion). By contrast, not having been trained as a professional epistemologist, Billy ignores such distant possibilities and infers that 'it is raining' is known.

Assuming that conditional reasoning underlies modal interpretation (cf. Section 4 on deriving the semantics from Korean conditional evaluatives), our theory of modality independently motivates such a felicity condition: our analysis of '*must* φ' involves reasoning with conditionals of the form 'if φ, then EVAL' as well as 'if ψ, then EVAL' for each alternative ψ to φ. It is well-known that an indicative conditional is felicitous only if its antecedent is a possibility (Stalnaker 1976). From our perspective, this implies that '*must* φ' is felicitous only if φ and each alternative to φ are possibilities. Insofar as some alternative to φ contains a $\neg\varphi$-world—which we believe to be a reasonable assumption—we cannot eliminate every $\neg\varphi$ possibility. As a consequence, epistemic necessity modals are felicitous only if the prejacent is not known.[38]

---

[38]Deontic necessity modals do not require such a felicity condition, and correctly so. In fact, Chung (2019) proposes that Korean conditional evaluatives receiving a deontic interpretation requires analyzing the conditional as a counterfactual conditional. If this is on the right track, our analysis of deontic modality will compare *causal* expected utilities as opposed to *evidential* expected utilities (Gibbard & Harper, 1978). An exploration of this subtle but substantive distinction is beyond the scope of the present article.

	Kratzer (2012)		Lassiter (2011)		Our theory	
	must	*ought*	*must*	*ought*	*must*	*ought*
Miners puzzle	✗	✗	✗	△	△	✓
Modal conjunction fallacy	✗	✗	✗	✗	✓	✓
Modal lawyers and engineers	✗	✗	✗	✗	✓	✓

Table 2 Predictions of theories of modality for the case studies discussed in detail in this article. For the miners puzzle, a cross (✗) indicates that the theory makes the wrong predictions; a triangle (△) appears in cases where the theory makes essentially the right predictions, with important caveats discussed in the main text in the appropriate sections; a checkmark (✓) indicates that the theory makes the right predictions. For the modal conjunction fallacy and modal lawyers and engineers, a cross means that the theory does not predict a fallacy, and checkmark means that the theory does.

8 Conclusion

This article presented a novel theory of modality in terms of comparisons between the expected values of the prejacent and its alternatives. We defined a general notion of "expected value" that allows for a single lexical entry to cash out expected value in terms of likelihoods as a proxy for explanatory value in the epistemic case, and in terms of expected utilities in the deontic case. The difference between the two cases, in our approach, lies purely in the properties of a contextually supplied set of propositions: facts in need of explanation in the epistemic case, ideals in the deontic case. Our proposal preserves the classical insight that very many languages of the world use a shared pool of modal constructions irrespective of modal domain, in that we give a single lexical entry for each modal operator that makes no distinction between the epistemic, deontic, or other modal domains. At the same time, our view incorporates the successes of more recent approaches to modality that avail themselves of the probability calculus and of decision-theoretic tools. We developed a detailed analysis of the strong necessity modal *must* in English and its Korean counterpart, a complex construction that we argue wears this kind of expected-value semantics on its sleeve. We also gave the beginnings of a semantics in the same spirit for weak necessity modals like *ought* or *should*, and we argued that an analysis of possibility modals in terms of duals of strong necessity in our system yields a reasonable interpretation for English *might* or *can*.

We considered three case studies in some detail, and evaluated the predictions of extant accounts of modality that are representative of the two main camps in the field: quantificational semantics based on ordinal relations between possible worlds, and probabilistic approaches. We summarize these predictions in Table 2.

This table is to be taken with a grain of salt. In particular, we are in no way claiming that other theories are constitutionally incapable of being modified in order to make the same predictions as our account. Regarding Kratzer's influential account, a central source of inspiration for our own version of a single lexical entry for each modal force and sophisticated modal backgrounds interacting interestingly to create different modal flavors, conjunction elimination for *must* is valid, making an account of our proposed modal conjunction fallacy extremely hard, if it is to be proposed within the

realm of modality itself. An articulated theory of modality and, say, representativeness à la Kahneman and Tversky (1973) is perhaps a reasonable way for this view to integrate our predictions, but such a combination is by no means a straightforward matter. Similar remarks apply to the quantificational approach in the case of our proposed modal lawyers and engineers puzzle, and the facts we summarize in the table for the miners puzzle are generally accepted in the field. On the probabilistic side, we find greater success with the miners puzzle, though more conservative predictions for the *must* case than our own, a matter that will likely require experimentation with naive participants to settle. For our novel epistemic puzzles on reasoning with *must*, extant probabilistic approaches, given their across-the-board adherence to Bayesian standards of rationality, make predictions entirely opposed to our theory's and, we have argued, to introspective intuitions.

Our somewhat radical new approach leaves multiple questions unanswered for the time being, beyond just whether our preliminary proposals for weak necessity and possibility modals are on the right track. In particular, in our effort to understand naive reasoning with epistemic *must* (a woefully understudied topic in the psychology of human reasoning), we could only sketch an analysis of how it is possible in a system like ours to still approximate the usual standards of rationality in terms of Bayesian update by factoring in prior probabilities via a plausibility requirement. Our project in this first article was to demonstrate with a detailed proof of concept the feasibility of our research program for modality, to show in particular that facts well established in linguistics and philosophy about the weakness of necessity modals in the epistemic case and similarly pervasive facts about apparent failures of human reasoning could be combined with a rational semantics in terms of expected utility for the deontic domain, all within one single lexical entry.

Declarations

References

Adams, E. (1965). The Logic of Conditionals. *Inquiry*, *8*(1-4), 166–197, https://doi.org/10.1080/00201746508601430

Ammann, A., & van der Auwera, J. (2002). Korean Modality - Asymmetries between Necessity and Possibility. *Selected Papers from the 12th International Conference on Korean Linguistics* (pp. 101–115).

Blumberg, K., & Hawthorne, J. (2022). Desire. *Philosophers' Imprint*, *22*(8), 1–17, https://doi.org/10.3998/phimp.2116

Blumberg, K., & Hawthorne, J. (2023). Inheritance: Professor Procrastinate and the logic of obligation. *Philosophy and Phenomenological Research*, *106*(1), 84–106, https://doi.org/10.1111/phpr.12846

Busemeyer, J.R., Pothos, E.M., Franco, R., Trueblood, J.S. (2011). A quantum theoretical explanation for probability judgment errors. *Psychological Review*, *118*(2), 193–218, https://doi.org/10.1037/a0022542

Busemeyer, J.R., Wang, Z., Pothos, E.M., Trueblood, J.S. (2015). The conjunction fallacy, confirmation, and quantum theory: Comment on Tentori, Crupi, and Russo (2013). *Journal of Experimental Psychology: General*, *144*(1), 236–243, https://doi.org/10.1037/xge0000035

Cariani, F. (2016a). Consequence and contrast in deontic semantics. *The Journal of Philosophy*, *113*(8), 396–416, https://doi.org/10.5840/jphil2016113826

Cariani, F. (2016b). Deontic Modals and Probability: One Theory to Rule Them All? N. Charlow & M. Chrisman (Eds.), *Deontic Modality* (pp. 11–46). Oxford University Press.

Cariani, F., Kaufmann, M., Kaufmann, S. (2013). Deliberative modality under epistemic uncertainty. *Linguistics and Philosophy*, *36*(3), 225–259, https://doi.org/10.1007/s10988-013-9134-4

Chung, W. (2019). Decomposing Deontic Modality: Evidence from Korean. *Journal of Semantics*, *36*(4), 665–700, https://doi.org/10.1093/jos/ffz016

Ciardelli, I., Zhang, L., Champollion, L. (2018). Two switches in the theory of counterfactuals. *Linguistics and Philosophy*, *41*(6), 577–621, https://doi.org/10.1007/s10988-018-9232-4

Crnič, L. (2011). *Getting even* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Crupi, V., Elia, F., Aprà, F., Tentori, K. (2018). Double conjunction fallacies in physicians' probability judgment. *Medical Decision Making*, *38*(6), 756–760, https://doi.org/10.1177/0272989X18786358

Crupi, V., Fitelson, B., Tentori, K. (2008). Probability, confirmation, and the conjunction fallacy. *Thinking & Reasoning*, *14*(2), 182–199, https://doi.org/10.1080/13546780701643406

Crupi, V., & Iacona, A. (2022). The Evidential Conditional. *Erkenntnis*, *87*(6), 2897–2921, https://doi.org/10.1007/s10670-020-00332-2

Deal, A.R. (2011). Modals without scales. *Language*, 559–585,

Douven, I. (2008). The evidential support theory of conditionals. *Synthese*, *164*(1), 19–44, https://doi.org/10.1007/s11229-007-9214-5

Dretske, F.I. (1972). Contrastive statements. *Philosophical Review*, *81*(4), 411–437, https://doi.org/doi.org/10.2307/2183886

Dulany, D.E., & Hilton, D.J. (1991). Conversational implicature, conscious representation, and the conjunction fallacy. *Social Cognition*, *9*(1), 85-110, https://doi.org/10.1521/soco.1991.9.1.85

Edwards, A.W.F. (1992). *Likelihood*. Johns Hopkins University Press.

Finlay, S. (2009). Oughts and Ends. *Philosophical Studies*, *143*(3), 315–340, https://doi.org/10.1007/s11098-008-9202-8

Fitelson, B. (1999). The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science*, *66*, 362–378, https://doi.org/10.1086/392738

Fitelson, B. (2007). Likelihoodism, Bayesianism, and relational confirmation. *Synthese*, *156*(3), 473–489, https://doi.org/10.1007/s11229-006-9134-9

Gaifman, H. (1988). A theory of higher order probabilities. B. Skyrms & W.L. Harper (Eds.), *Causation, chance and credence* (Vol. 41). Springer, Dordrecht.

Giannakidou, A., & Mari, A. (2016). Epistemic future and epistemic *MUST*: Non-veridicality, evidence, and partial knowledge. J. Błaszczak, A. Giannakidou, D. Klimek-Jankowska, & K. Migdalski (Eds.), *Mood, aspect, modality revisited: New answers to old questions* (pp. 75–117). University of Chicago Press.

Gibbard, A. (1980). Two recent theories of conditionals. *Ifs* (pp. 211–247). Springer.

Gibbard, A., & Harper, W.L. (1978). Counterfactuals and two kinds of expected utility. W.L. Harper, R. Stalnaker, & G. Pearce (Eds.), *Foundations and applications of decision theory* (pp. 125–162). Springer.

Goble, L. (1996). Utilitarian deontic logic. *Philosophical Studies*, *82*(3), 317–357, https://doi.org/10.1007/BF00355312

Goldstein, S., & Santorio, P. (2021). Probability of epistemic modalities. *Philosophers' Imprint*, *21*(33), 1–37,

Goodhue, D. (2017). *Must φ is felicitous only if φ is not known*. *Semantics and Pragmatics*, *10*(14), 1–27, https://doi.org/10.3765/sp.10.14

Grice, P. (1975). Logic and conversation. P. Cole & J.L. Morgan (Eds.), *Syntax and semantics: Speech acts* (Vol. 3, pp. 41–58). New York: Academic Press.

Guerrini, J., Sablé-Meyer, M., Mascarenhas, S. (2022). *An explanation of representativeness: Contrastive confirmation-theoretical reasoning motivated by question-answer dynamics.* (Talk given at the 44th Annual Meeting of the Cognitive Science Society)

Heim, I. (1992). Presupposition Projection and the Semantics of Attitude Verbs. *Journal of semantics*, *9*(3), 183–221, https://doi.org/10.1093/jos/9.3.183

Jackson, F. (1979). On assertion and indicative conditionals. *The Philosophical Review*, *88*(4), 565–589, https://doi.org/10.2307/2184845

Jackson, F. (1985). On the semantics and logic of obligation. *Mind*, *94*(374), 177–195, https://doi.org/10.1093/mind/XCIV.374.177

Jackson, F., & Pargetter, R. (1986). Oughts, Options, and Actualism. *The Philosophical Review*, *95*(2), 233–255, https://doi.org/10.2307/2185591

Jeffrey, R.C. (1965). *The Logic of Decision*. University of Chicago Press.

Jeffrey, R.C., & Edgington, D. (1991). Matter-of-fact conditionals. *Proceedings of the Aristotelian Society, Supplementary Volumes*, *65*, 161–209,

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological review*, *80*(4), 237–251, https://doi.org/10.1037/h0034747

Katzir, R. (2007). Structurally-defined alternatives. *Linguistics and Philosophy*, *30*, 669–690, https://doi.org/10.1007/s10988-008-9029-y

Kaufmann, M. (2017). What 'may' and 'must' may be in Japanese. K. Funakoshi, S. Kawahara, & C. Tancredi (Eds.), *Japanese/Korean Linguistics 24.*

Kaufmann, S. (2005). Conditional predictions. *Linguistics and Philosophy*, *28*(2), 181–231, https://doi.org/10.1007/s10988-005-3731-9

Kaufmann, S. (2013). Causal premise semantics. *Cognitive Science*, *37*(6), 1136–1170, https://doi.org/10.1111/cogs.12063

Kaufmann, S. (2017). The limit assumption. *Semantics and Pragmatics*, *10*(18), 1–29, https://doi.org/10.3765/sp.10.18

Kennedy, C. (2007). Vagueness and grammar: the semantics of relative and absolute gradable adjectives. *Linguistics and philosophy*, *30*(1), 1–45, https://doi.org/10.1007/s10988-006-9008-0

Kennedy, C., & McNally, L. (2005). Scale Structure, Degree Modification, and the Semantics of Gradable Predicates. *Language*, 345–381,

Klecha, P. (2014). *Bridging the divide: Scalarity and modality* (Unpublished doctoral dissertation). The University of Chicago.

Koehler, J.J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, *19*(1), 1–17, https://

doi.org/10.1017/S0140525X00041157

Kolodny, N., & MacFarlane, J. (2010). Ifs and oughts. *The Journal of philosophy*, *107*(3), 115–143,

Kratzer, A. (1979). Conditional necessity and possibility. R. Bäuerle, U. Egli, & A. von Stechow (Eds.), *Semantics from different points of view* (pp. 117–147). Springer.

Kratzer, A. (1981). The Notional Category of Modality. H.J. Eikmeyer & H. Rieser (Eds.), *Words, Worlds, and Contexts, New Approaches to Word Semantics* (pp. 38–74). Berlin: Walter de Gruyter.

Kratzer, A. (1991). Modality. A. von Stechow & D. Wunderlich (Eds.), *Semantics: An international handbook of contemporary research* (pp. 639–650). De Gruyter.

Kratzer, A. (2012). *Modals and conditionals: New and revised perspectives* (Vol. 36). Oxford University Press.

Kyburg, H.E. (1961). *Probability and the logic of rational belief*. Wesleyan University Press.

Lassiter, D. (2011). *Measurement and modality: The scalar basis of modal semantics* (Ph.D. thesis). New York University.

Lassiter, D. (2017). *Graded modality: Qualitative and quantitative perspectives*. Oxford University Press.

Lewis, D. (1973). *Counterfactuals*. Oxford: Oxford University Press.

Lewis, D. (1976). Probabilities of Conditionals and Conditional Probabilities. W.L. Harper, R. Stalnaker, & G. Pearce (Eds.), *Ifs* (pp. 129–147). Springer.

Lewis, D. (1996). Elusive knowledge. *Australasian Journal of Philosophy*, *74*(4), 549–567, https://doi.org/10.1080/00048409612347521

Mangiarulo, M., Pighin, S., Polonio, L., Tentori, K. (2021). The effect of evidential impact on perceptual probabilistic judgments. *Cognitive Science*, , https://doi.org/10.1111/cogs.12919

McGee, V. (1985). A counterexample to modus ponens. *The Journal of Philosophy*, *82*(9), 462–471, https://doi.org/10.2307/2026276

Meyer, M.-C. (2013). *Ignorance and grammar* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Morier, D.M., & Borgida, E. (1984). The Conjunction Fallacy: A Task Specific Phenomenon? *Personality and Social Psychology Bulletin*, *10*(2), 243–252, https://doi.org/10.1177/0146167284102

Ofir, C. (1988). Pseudodiagnosticity in judgment under uncertainty. *Organizational Behavior and Human Decision Processes*, *42*(3), 343–363, https://doi.org/10.1016/0749-5978(88)90005-2

Pasternak, R. (2016). A conservative theory of gradable modality. *Semantics and Linguistic Theory 26* (pp. 371–390).

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Pearl, J. (2013). Structural Counterfactuals: A Brief Introduction. *Cognitive Science*, *37*(6), 977–985, https://doi.org/10.1111/cogs.12065

Sablé-Meyer, M., & Mascarenhas, S. (2022). Indirect illusory inferences from disjunction: a new bridge between deductive inference and representativeness. *Review of Philosophy and Psychology*, *13*, 567–592, https://doi.org/10.1007/s13164-021-00543-8

Santorio, P. (2019). Interventions in premise semantics. *Philosophers' Imprint*, *19*(1), 1–27,

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy*, *27*, 367-391, https://doi.org/10.1023/B:LING.0000023378.71748.db

Schulz, K. (2011). If you'd wiggled A, then B would've changed. *Synthese*, *179*(2), 239–251, https://doi.org/10.1007/s11229-010-9780-9

Sloman, A. (1970). 'Ought' and 'Better'. *Mind*, *79*(315), 385–394, https://doi.org/10.1093/mind/lxxix.315.385

Stalnaker, R.C. (1976). Indicative conditionals. W.L. Harper, R. Stalnaker, & G. Pearce (Eds.), *Ifs* (pp. 193–210). Springer.

Stalnaker, R.C. (1978). Assertion. P. Cole (Ed.), *Pragmatics* (Vol. 9, pp. 315–332). New York: Academic Press.

Stolarz-Fantino, S., Fantino, E., Zizzo, D.J., Wen, J. (2003). The conjunction effect: New evidence for robustness. *American Journal of Psychology*, *116*(1), 15–34, https://doi.org/10.2307/1423333

Tentori, K., Crupi, V., Russo, S. (2013). On the determinants of the conjunction fallacy: Probability versus inductive confirmation. *Journal of Experimental Psychology: General*, *142*(1), 235–255, https://doi.org/10.1037/a0028770

Tversky, A., & Kahneman, D. (1983). Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review*, *90*(4), 293–315, https://doi.org/10.1037/0033-295X.90.4.293

Villalta, E. (2008). Mood and gradability: an investigation of the subjunctive mood in Spanish. *Linguistics and Philosophy*, *31*, 467-522, https://doi.org/10.1007/s10988-008-9046-x

von Fintel, K., & Gillies, A.S. (2010). *Must... stay... strong! Natural language semantics*, *18*(4), 351–383, https://doi.org/10.1007/s11050-010-9058-2

von Fintel, K., & Iatridou, S. (2005). *What to do if you want to go to Harlem: Anankastic conditionals and related matters* (Unpublished manuscript). MIT.

Wells, G.L., & Harvey, J.H. (1977). Do people use consensus information in making causal attributions? *Journal of Personality and Social Psychology*, *35*(5), 279–293, https://doi.org/10.1037/0022-3514.35.5.279

Yalcin, S. (2005). A puzzle about epistemic modals. *MIT Working Papers in Linguistics*, *51*, 231–272,

Yalcin, S. (2010). Probability Operators. *Philosophy Compass*, *5*(11), 916–937, https://doi.org/https://doi.org/10.1111/j.1747-9991.2010.00360.x