Modality, expected utility, and hypothesis testing

WooJin Chung and Salvador Mascarenhas
Institut Jean Nicod, Département d'études cognitives,
ENS, EHESS, CNRS, PSL Research University

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.

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.

Informally, a sentence 'must φ' will be true just in case assuming φ would lead to a greater *expected value* than any of the 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 are not about maximizing posterior probability of the prejacent, conditional on some epistemic facts. Instead, they are about maximizing the explanatory power of the prejacent as a predictor of contextually relevant epistemic facts. 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 $Pr(e \mid \varphi)$ is sufficiently higher than $Pr(e \mid \psi)$, for each ψ that is a relevant alternative to φ.

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, allowing us for example to provide an account of the miners puzzle (Kolodny and 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 offers a good-enough explanation for a contextually determined, salient body of evidence. Accordingly, we immediately account for the evidential flavor of epistemic *must*.

More interestingly, 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 and Kahneman, 1983), participants read a description of an individual named Linda that asserts that in her youth she engaged with political activism. Participants are then 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 "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 just in case 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 et al., 2008).

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 and 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.'
'Jack must/should 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* (Kaufmann, 2017a) 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* 'EVAL' with the conditional semantics to derive our theory of modality.

1.1 Two theories of modality

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

The classical theory due to Angelika Kratzer is a quantification-based approach. Assuming the Limit Assumption (Lewis, 1973; Kaufmann, 2017b), 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. 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 establishess an arithmetic relationship between degrees:

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

Prior to Lassiter, Yalcin (2010) observed that extant theories of comparative modality based on qualitative ordering validates 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, all necessity modals require that the *expected utility* of the prejacent be greater than some threshold θ. Just like weak epistemic necessity, the threshold θ of deontic *should* or *ought* is sensitive to contextually salient alternatives.

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 explains better the distribution of degree modifiers. However, the innovation comes at the cost of ignoring the cross-linguistic generalization that modals 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 (modulo the selection of conversational backgrounds). 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.[1]

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).

(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

[1] We would also like to note that one cannot reduce the probability weights in an expected utility formula (e.g., (9)) 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.

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 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.

$$D(h, e) = Pr(h|e) - Pr(h)$$

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.

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

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 presently. But we will also illustrate how a more properly Bayesian measure can be achieved.

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/should/ought 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 (i.e., pieces of evidence) for epistemics.

To formalize our proposal, we first define $\mathbb{E}[\psi \mid \varphi]$ as in (5). 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

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

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. For epistemics, the explanatory value of φ is greater than θ but no alternative to φ is such that its explanatory value is greater than θ.[2] When a salient set of alternatives is not given, the alternative set consists of the prejacent φ and its negation $\neg\varphi$.

(6) Proposal

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

We find it more intuitive to read the formula as follows: In a deontic context, φ is *the only good choice* among the available options. In an epistemic context, φ is *the only good explanation* of the evidence among the salient 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) $\mu_{\text{EVAL}} = \lambda w. \mid \{r \in R \mid r \text{ is true at } w\} \mid,$
where R 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 (i.e., 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. For example, if all ideals are realized and all rules are abided by at w, then w will be a better world than any other world. In this sense, the number of ideals/rules realized/abided by at a given world is the *utility value* of the world. Thus,

[2]We will remain agnostic as to how strong necessity modals (e.g., *must*) differ from weak necessity modals (e.g., *should* and *ought*).

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}

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

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:

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

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

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.[3]

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

(10) Epistemic interpretation of μ_{EVAL}

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

For the epistemic interpretation of $\mathbb{E}[\mu_{\text{EVAL}} \mid \varphi]$, 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 $\mathbb{E}[\mu_{\text{EVAL}} \mid \varphi]$ 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 $e_i \in R_E$.

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

$$\begin{aligned}\mu_{\text{EVAL}} &= \lambda w. \mid \{e \in R_E \mid e \text{ is true at } w\} \mid \\ &= \lambda w. \sum_{i=1}^{n} e_i(w), \quad \text{where } R_E = \{e_1, ..., e_n\}\end{aligned}$$

[3] 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) Epistemic measure: $\mathbb{E}[\mu_{\text{EVAL}} \mid \varphi]$ as the explanatory value of φ

$$\begin{aligned}\mathbb{E}_w[\mu_{\text{EVAL}} \mid \varphi] &= \sum_{w_j \in \varphi} \mu_{\text{EVAL}}(w_j) \times Pr_w(\{w_j\} \mid \varphi) \\ &= \sum_{w_j \in \varphi} \sum_{i=1}^{n} e_i(w_j) \times Pr_w(\{w_j\} \mid \varphi) \\ &= \sum_{i=1}^{n} \sum_{w_j \in \varphi} e_i(w_j) \times Pr_w(\{w_j\} \mid \varphi) \\ &= \sum_{i=1}^{n} Pr_w(e_i \mid \varphi), \quad \text{where } R_E = \{e_1, ..., e_n\}\end{aligned}$$

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 numerically represent the degree to which φ supports given pieces of evidence, we call this measure the *explanatory value* of φ.

The analysis of epistemic modality is what crucially distinguishes our theory from Lassiter's theory. Lassiter claims that epistemic modals compare the *(posterior)* probability of the prejacent to a contextually determined threshold, whereas we propose that epistemic modals concern the *explanatory value* of φ which is based on likelihoods. This feature will allow our theory to capture inferential uses of modality and naive inferences such as the conjunction fallacy.

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. We would like to point out that an alternative formulation (though not entirely equivalent) that makes *direct* comparisons is closely related to the L confirmation measure discussed in the introduction. 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

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

If we assume that (i) the only alternative to φ is its negation, (ii) there is a single piece of evidence,[4] and (iii) take the logarithm of each measured value, 'ought φ' is true if and only if $L(\varphi, e)$ is greater than the contextually supplied threshold θ, as shown below:[5]

[4]In fact, it is widely assumed in the literature that several pieces of evidence are combined into a single piece of evidence. In our terms, this amounts to conjoining the relevant known facts $e_i \in R_E$ and using the conjunction as evidence. While it is a common practice in Bayesian confirmation theory to do so, we leave it open as to whether this process is a necessary part of the interpretation of modals.

[5]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 $Alt(\varphi) = \{\neg\varphi\}$)

$$\llbracket \text{must/should/ought}_{\text{LOG,ALT}}\ \varphi \rrbracket^w$$
$$= \forall\psi \in Alt(\varphi) : log(\mathbb{E}_w[\mu_{\text{EVAL}} \mid \varphi]) - log(\mathbb{E}_w[\mu_{\text{EVAL}} \mid \psi]) > \theta$$
$$= log(\mathbb{E}_w[\mu_{\text{EVAL}} \mid \varphi]) - log(\mathbb{E}_w[\mu_{\text{EVAL}} \mid \neg\varphi]) > \theta$$
$$= log(Pr_w(e \mid \varphi)) - log(Pr_w(e \mid \neg\varphi)) > \theta$$
$$= log\left(\frac{Pr_w(e \mid \varphi)}{Pr_w(e \mid \neg\varphi)}\right) > \theta$$
$$= L(\varphi, e) > \theta$$

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 is the source of irrationality in assessing comparative subjective probabilities with epistemics. Note that expected utility is a rational measure employed in decision theory. By contrast, explanatory value—sum of likelihoods—is a measure that gives rise to fallacious conclusions in reasoning tasks (Crupi et al., 2008). Also, as we have seen above, the modal semantics that uses explanatory values effectively makes reference to the L confirmation measure, which is a measure of inductive confirmation. In the following section, we will see that two puzzles in the psychology of reasoning, the lawyers and engineers puzzle and the conjunction fallacy, are naturally accounted for. Moreover, the very same semantics (modulo the modal flavor) offers a decision theoretic analysis of the miners puzzle. While various proposals have been made for the lawyers and engineers puzzle and the conjunction fallacy, our theory remains unique in two respects: (i) it receives independent support from research on modality in linguistics and philosophy, and (ii) it explains why people opt for an irrational strategy despite the existence of a rational alternative: reasoning about comparative subjective probabilities is a reasoning strategy used by humans, and this is reflected in the semantics of modals.

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 and 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 and Tversky, 1973; Tversky and Kahneman, 1983).

3.1 The miners puzzle (Kolodny and MacFarlane, 2010)

As Lassiter (2011) points out, an expected utility-based 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 and MacFarlane, 2010). Given the situation described in Table 1, examples (15a)–(15c) are all intuitively true.

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

(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 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 brief summary of proof:[6]

(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)-(20) cannot all be true.

Kolodny and MacFarlane argue that the issue arises because Kratzer's conversational back-

[6]The readers are referred to Cariani et al. (2013) for a more detailed discussion concerning all possible interpreations of the miners examples.

grounds are not seriously information-dependent, that is, one's preference cannot change upon obtaining new information.

An expected utility-based analysis of the miners puzzle naturally encodes this information dependence into the semantics, as conditioning on new information adjusts the probability weights used to calculate expected utilities. In what follows, we demonstrate that this is indeed the case.

Regarding 'we ought to block neither shaft', the gist of our analysis is that the expected utility of blocking neither shaft (i.e., **block-neither**) is greater than the expected utility of blocking shaft A (i.e., **block-A**) or 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):

(17) $R_D = \{$ 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, since the setup is agnostic as to their exact whereabouts. 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 (18).

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

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

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 (19), 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 action and the miners' whereabouts are independent. Analogously, the expected utility of **block-B** is also 5.

(19) The expected utility of ‘we block shaft A’

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

(20) The expected utility of ‘we block shaft B’

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

We analyze (15a) as in (21). Informally, “blocking neither shaft is the only good choice among the available options”. This is true since **block-neither** is appreciably better than **block-A** and **block-B**.

(21) $[\![$ ought **block-neither**$]\!]^w$

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

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 condition on the antecedent proposition.[7]

Conditionalizing on **miners-in-A** does not change the expected utility of **block-neither** since

[7] 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** $\wedge$ **block-A**, then EVAL/suffice’ if we take the Import-Export Principle (Gibbard, 1981; McGee, 1985) for granted. Given the assumptions to be presented in section 4, the latter denotes the expected utility of **miners-in-A** $\wedge$ **block-A**.

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 (22). 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**.

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

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

We flesh out our analysis of (15b) in (23). Informally, 'given that the miners are in shaft A, blocking shaft A is the only good choice among the available options'.

(23) $[\![$ if **miners-in-A**, ought **block-A** $]\!]^w$

$$\begin{aligned}
&= \mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{miners-in-A} \wedge \textbf{block-A}] > \theta \\
&\quad \wedge \mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{miners-in-A} \wedge \textbf{block-neither}] \leq \theta \\
&\quad \wedge \mathbb{E}_w[\mu_{\text{EVAL}} \mid \textbf{miners-in-A} \wedge \textbf{block-B}] \leq \theta
\end{aligned}$$

What we presented in this section is more or less a reproduction of Lassiter's analysis. This is no surprise because both theories compare expected utilities of contextually salient alternatives. Our contribution on the deontic side is that we *compositionally derive* the expected utility-based semantics from natural language data, which we present in Section 4.

3.2 The conjunction fallacy (Tversky and Kahneman, 1983)

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

(24) 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, if faced with the same setup as the classical task but the following two competing options to choose from:

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

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 (25b) is in a clear sense more attractive than (25a). 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.

If (25) 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. Specifically, our analysis of the modal sentences in (25) reduces to a comparison between the explanatory values of (24a) and (24b), with the description of Linda as the relevant evidence. As Crupi et al. (2008) show, such an account predicts that reasoners should find (25b) more attractive than (25a).

3.3 Lawyers and engineers (Kahneman and 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 (26). Kahneman and Tversky tested two conditions between participants: in one, Jack's description was drawn randomly from a sample of 70 lawyers and 30 engineers, in the other the sample consisted instead of 30 lawyers and 70 engineers. They found that participants' judgments were unaffected by these prior probabilities: they consistently assigned a high probability to Jack's being an engineer. This suggests that indeed they were not resorting to the normative standard provided by Bayes' theorem to decide on their response.

(26) 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.*

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:

(27) Jack must/should/ought to be an engineer.

This is just as surprising as the reported result in the original experiment: to assent to (27) in such an experimennt 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.

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.[8]

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:[9] [10]

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

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

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.

(30) The explanatory value of Jack being an engineer

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

[8] 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.

[9] The probabilities were taken from Janek Guerrini’s experiments on the lawyers and engineers scenario.

[10] We would like to note that 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 mseasure 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.

(31) The explanatory value of Jack being a lawyer

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

Given the above probability assignments, 'ought **engineer**' is true if and only if 'the hypothesis that Jack is an engineer is the good explanation of the given evidence among the hypotheses'.

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

4 Natural language evidence: conditional evaluatives

In this section, we compositionally derive our proposed semantics from Korean conditional evaluatives (repeated below as (33)), which have a transparent morphosyntax.

(33) 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.'
'Jack must/should be home.'

We conjecture that the above conditional evaluative construction is the transparent version of English necessity modal expressions (e.g., *must*, *should*, *ought*). 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.[11] 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.[12]

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.

[11] 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 s/he has a different understanding of obligation as a non-native speaker of English?

[12] 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.

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 (34).[13]

(34) $[\![\text{ EVAL }]\!]^w = \mu_{\text{EVAL}} = \lambda w'. \mid \{r \in R \mid r(w') = 1\} \mid,$
where R is the set of relevant propositions at w

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 ψ' is the expected value of ψ given φ.[14]

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

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 the proposition ψ given φ is by definition the conditional probability of ψ given φ. This proves that 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 Lewis, 1976; Jackson, 1979; Gibbard, 1981; Jeffrey and Edgington, 1991; Kaufmann, 2005; Crupi and Iacona, 2019, for relevant work in linguistic and philosophy). However, we depart from the 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 ψ of the conditional in (35) 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.

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

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 thresh-

[13]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.'

[14]We want to make it clear that we are not claiming that (35) 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 (2019).

olding operation.[15]

(37) 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 (38) 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.

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

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.[16] This is exactly what we proposed for the analysis of modal necessity in (6).

(39) The compositional semantics of Korean conditional evaluatives

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

Hence we have independence 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 allows our theory to address puzzles of failure of reasoning (i.e., why do people make fallacious inferences?), it naturally raises a question as to how the theory can explain the puzzle of success (i.e., how

[15]The thresholding operation is reminiscent of the *pos* morpheme in 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.

[16]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.

can people make classically sound inferences despite all?), since priors or comparable qualitative orderings are precisely what a theory needs to take into account to address the issue.

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 certain 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 problem 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 (40). Upon hearing that John did not come to work, one could reasonably conjecture that he 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.

(40) a. John did not come to work today.
b. He must have caught a cold.
c. #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 him 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 theory incorrectly predicts that (40b) is false whereas (40c) is true. Note that the prediction remains unaltered even if one entertains a different alternative to **dead** such as 'John caught a cold', as $Pr(\textbf{absent} \mid \textbf{dead}) \gg Pr(\textbf{absent} \mid \textbf{cold})$.

(41) Likelihood-based comparison

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

(42) 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. 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:

(43) Plausible probability assignments

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

(44) Corresponding D measures

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

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 two 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. Since this connection remains at the heart of our theory, we must seek alternative routes to account for the sensitivity to priors.

One possible way to capture the contrast in (40) 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.[17] In this view, (40b) is a reasonable thing to say because a cold is quite common a condition and accordingly has a relatively high prior. By contrast, (40c) is odd 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.

Note also that 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 (Wells and Harvey, 1977; Ofir, 1988; Koehler, 1996).

One could introduce an alternative requirement: that the prior probability of the modal prejacent be comparable to those of its alternatives. In a typical context, we will have that $Pr(\textbf{dead}) \ll Pr(\neg\textbf{dead})$, so it is odd to say 'John must be dead' as opposed to not dead. Similarly for the alternative set $\{\textbf{dead}, \textbf{cold}\}$. Granting that epistemic modal reasoning involves hypothesis testing, this comparability requirement can be understood as a necessary condition for the prejacent to qualify as a *competing* hypothesis.

Lastly, one could posit a novel measure that takes into account both the explanatory value

[17] We speculate additionally that this carries over interestingly to the deontic domain, with connections with plausibility (and possibly doability) of deontic injunctions.

and the prior. The simplest implementation of this idea is to calculate the weighted sum of the explanatory value and the prior:

(45) Weighted sum of the explanatory value and the prior of φ

$$\alpha \times \mathbb{E}_w[\mu_{\text{EVAL},w} \mid \varphi] + (1-\alpha) \times Pr_w(\varphi)$$

Our suggestions are only tentative and we do not mean to present in this article a comprehensive theory that incorporates both explanatory values and priors. We leave it to future research to reveal the exact role of priors in modal semantics and modal reasoning.

6 Further implication: the weakness of epistemic necessity

Sentences with epistemic 'must' are deviant when the prejacent is known. Some have argued for a felicity condition on 'must' utterances to this very effect (Giannakidou and Mari, 2016; Goodhue, 2017). Others have argued that this effect is a consequence of a different felicity condition: for von Fintel and Gillies (2010) 'must' utterances require that the evidence that justifies a conclusion of the prejacent be indirect.

The account we offered here easily cashes out the idea of 'must' as requiring indirect evidence. We proposed that epistemic "must φ" asserts that φ is the best explanation for a contextually determined, salient body of evidence. It stands to reason that φ cannot felicitously be a direct transplant from the relevant body of evidence: a piece of evidence is unacceptable as an explanation of itself.

Additionally, our proposal also cashes out the idea that any epistemic 'must' sentence with a known prejacent should be infelicitous, above and beyond the natural requirement of non-trivial explanations we just discussed.

Goodhue (2017) explains the contrast between (46) and (47) in terms of whether the prejacent 'it is raining' is known to the speaker. He argues that (46) is infelicitous because Billy knows that it is raining, whereas (47) is felicitous because she cannot infer that it is raining from observing someone holding a wet umbrella.

(46) *Billy is looking out the window at the pouring rain.*
It must be raining.

(47) *Billy sees someone enter the building holding a wet umbrella, but she cannot see outside herself.*
It must be raining.

What strengthens Goodhue's argument is that from the perspective of a skeptical epistemologist, 'it must be raining' can be felicitous even when she observes the pouring rain, as in (48).

(48) *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:

(49) Lewis (1996) on knowledge
The speaker knows that φ
$\leftrightarrow$ 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.

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. From our perspective, this implies that 'must φ' is felicitous only if φ and each alternative to φ are a possibility. Insofar as some alternative to φ is 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 felicitously only if the prejacent is not known.[18]

7 Conclusion

This paper presents a novel theory of modality where its interpretation involves comparing the expected values of the prejacent and its alternatives. We revive one key insight of the classical theory of modality within a probability-based framework, namely that modal flavor is solely determined by context. In accordance with this view, expected value calculation yields different measures depending on the contextually supplied body of information. When interpreted deontically, expected value calculation yields the decision-theoretic conception of expected utility. When interpreted epistemically, expected value calculation measures the explanatory power of the prejacent as a predictor of contextually relevant epistemic facts.

Reconciling the two types of modals is not only theoretically preferable but also has interesting consequences. In addition to preserving the decision-theoretic conception of deontic modality via expected utility, our theory makes novel predictions in the epistemic domain. We give an immediate account of modal variants of the conjunction fallacy and the lawyers and engineers puzzle, demonstrating that reasoning problems from the heuristics and biases literature are subject to linguistic and philosophical analysis. Our theory suggests that naive human reasoners are not blind to probability, but rather they are applying a strategy that could have been a rational one

[18]One could point out that 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 and Harper, 1978).

(e.g., when applied to decision making) to the wrong domain. This erroneous choice of strategy is induced by people's understanding of modality that deontic and epistemic modals are interpreted in the same way, modulo the body of information attended to.

Lastly, our theory receives independent support from natural language data. We derive the modal semantics in an entirely transparent manner from the Korean necessity modal construction, which consists of a conditional and an evaluative predicate. Language data have proved useful in the study of human cognition when coupled with various methodologies. This work additionally shows that the study of linguistic modality provides insights into human reasoning.

References

Adams, Ernest. 1965. The Logic of Conditionals. *Inquiry* 8:166–197.

Ammann, Andreas, and Johan van der Auwera. 2002. Korean Modality - Asymmetries between Necessity and Possibility. In *Selected Papers from the 12th International Conference on Korean Linguistics*, 101–115.

Cariani, Fabrizio, Magdalena Kaufmann, and Stefan Kaufmann. 2013. Deliberative modality under epistemic uncertainty. *Linguistics and philosophy* 36:225–259.

Chung, WooJin. 2019. Decomposing Deontic Modality: Evidence from Korean. *Journal of Semantics* 36:665–700.

Crupi, Vincenzo, Branden Fitelson, and Katya Tentori. 2008. Probability, confirmation, and the conjunction fallacy. *Thinking & Reasoning* 14:182–199. Publisher: Taylor & Francis.

Crupi, Vincenzo, and Andrea Iacona. 2019. The evidential conditional. *Unpublished manuscript* .

Deal, Amy Rose. 2011. Modals without scales. *Language* 559–585.

Douven, Igor. 2008. The evidential support theory of conditionals. *Synthese* 164:19–44. Publisher: Springer.

Edwards, Anthony William Fairbank. 1992. *Likelihood*. Johns Hopkins University Press.

von Fintel, Kai, and Anthony S Gillies. 2010. *Must*... stay... strong! *Natural language semantics* 18:351–383. Publisher: Springer.

Fitelson, Branden. 1999. The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science* 66:362–378.

Fitelson, Branden. 2007. Likelihoodism, Bayesianism, and relational confirmation. *Synthese* 156:473–489.

Giannakidou, Anastasia, and Alda Mari. 2016. Epistemic future and epistemic *must*: nonveridicality, evidence, and partial knowledge. In *Mood, aspect, modality revisited: New answers to old questions*, ed. Joanna Błaszczak, Anastasia Giannakidou, Dorota Klimek-Jankowska, and Krzysztof Migdalski, 75–117. Chicago, IL: University of Chicago Press.

Gibbard, Allan. 1981. Two recent theories of conditionals. In *Ifs*, 211–247. Springer.

Gibbard, Allan, and William L Harper. 1978. Counterfactuals and two kinds of expected utility. In *Foundations and applications of decision theory*, 125–162. Springer.

Goodhue, Daniel. 2017. *Must* φ is felicitous only if φ is not known. *Semantics and Pragmatics* 10:14.

Jackson, Frank. 1979. On assertion and indicative conditionals. *The Philosophical Review* 88:565–589.

Jeffrey, Richard C. 1965. *The Logic of Decision*. Chicago, IL: University of Chicago Press.

Jeffrey, Richard C., and Dorothy Edgington. 1991. Matter-of-fact conditionals. *Proceedings of the Aristotelian Society, Supplementary Volumes* 65:161–209.

Kahneman, Daniel, and Amos Tversky. 1973. On the psychology of prediction. *Psychological review* 80:237–251.

Kaufmann, Magdalena. 2017a. What 'may' and 'must' may be in Japanese. In *Japanese/Korean Linguistics 24*, ed. Kenshi Funakoshi, Shigeto Kawahara, and Chris Tancredi.

Kaufmann, Stefan. 2005. Conditional predictions. *Linguistics and Philosophy* 28:181–231.

Kaufmann, Stefan. 2017b. The limit assumption. *Semantics and Pragmatics* 10.

Kennedy, Christopher. 2007. Vagueness and grammar: the semantics of relative and absolute gradable adjectives. *Linguistics and philosophy* 30:1–45.

Kennedy, Christopher, and Louise McNally. 2005. Scale Structure, Degree Modification, and the Semantics of Gradable Predicates. *Language* 345–381.

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

Kolodny, Niko, and John MacFarlane. 2010. Ifs and oughts. *The Journal of philosophy* 107:115–143.

Kratzer, Angelika. 1981. The Notional Category of Modality. In *Words, Worlds, and Contexts, New Approaches to Word Semantics*, ed. H.-J Eikmeyer and H. Rieser, 38–74. Berlin: Walter de Gruyter.

Kratzer, Angelika. 1991. Modality. In *Semantics: An international handbook of contemporary research*, ed. A. von Stechow and D. Wunderlich, 639–650. Berlin: de Gruyter.

Kratzer, Angelika. 2012. *Modals and conditionals: New and revised perspectives*, volume 36. Oxford University Press.

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

Lassiter, Daniel. 2017. *Graded modality: Qualitative and quantitative perspectives*. Oxford University Press.

Lewis, David. 1973. *Counterfactuals*. Oxford: Oxford University Press.

Lewis, David. 1976. Probabilities of Conditionals and Conditional Probabilities. In *Ifs*, 129–147. Springer.

Lewis, David. 1996. Elusive knowledge. *Australasian Journal of Philosophy* 74:549–567.

McGee, Vann. 1985. A counterexample to modus ponens. *The Journal of Philosophy* 82:462–471.

Ofir, Chezy. 1988. Pseudodiagnosticity in judgment under uncertainty. *Organizational Behavior and Human Decision Processes* 42:343–363.

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge, UK: Cambridge University Press.

Pearl, Judea. 2013. Structural Counterfactuals: A Brief Introduction. *Cognitive Science* 37:977–985.

Tentori, Katya, Vincenzo Crupi, and Selena Russo. 2013. On the determinants of the conjunction fallacy: Probability versus inductive confirmation. *Journal of Experimental Psychology: General* 142:235.

Tversky, Amos, and Daniel Kahneman. 1983. Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review* 90:293–315.

Wells, Gary L, and John H Harvey. 1977. Do people use consensus information in making causal attributions? *Journal of Personality and Social Psychology* 35:279.

Yalcin, Seth. 2010. Probability Operators. *Philosophy Compass* 5:916–937.