Deontic Modals and Probabilities: One Theory to Rule Them All?

Fabrizio Cariani, Northwestern University (`cariani.org`)

[*Definition:* ☕ := "Here be complexities. Talk about them over coffee"]

1. Aim

Central Aim: to design a probabilistic premise semantics for deontic *should* (and *ought*).

Specifically, a semantic account that mixes probabilistic structure with the flexibility of classical, Kratzer-style premise semantics.

The resulting theory has more structure than classical premise semantics (by incorporating probabilities)...

... but less structure than scalar theories (Lassiter 2011), by not incorporating numerical value functions or expectational structure.

Main message: the insight that the meaning of *should*/*ought* involves probability should not be identified with the claim that their meaning is tied to expectation.

2. Motivation

Why add probabilistic structure in a semantics for *should*?

Assumptions:

[A1] ⌜ likely φ⌝ constrains a probabilistic information state.
[A2] ⌜ if it's likely that φ ...⌝ shifts the information state.

Judgments (Yalcin, 2012a):

(1) a. If it's likely that you will fail the class, you should drop it.
b. If it's not likely that you will fail the class, you should not drop it.

▶ *Claim:* if A1/A2 are accepted, we need an analysis of *should* that makes it sensitive to the shift introduced by *likely*.

Note: I know some accounts of *likely* violate A1/A2. I am not out to refute them. This paper is based on *assuming* A1/A2: other starting points are possible. (☕).

▶ *Related Judgment:* (1a) together with "it is twice as likely that you will fail the class than not" entails "you should drop the class". But Lassiter (2014) notes that "twice as likely" is best made sense of on a model that satisfies A1.

3. Related Issue: Information Sensitivity

▶ Kolodny & MacFarlane (2010) argue that semantic values of deontic claims depend on background information states (motivated by examples like (2)-(3) below).

▶ I agree with this program and go a bit further: sensitivity to probabilistic states is injected in the semantics by the *same mechanism* that yields information sensitivity. (☕) The semantics in §4 is a probabilistic generalization of the information-sensitive, but non-probabilistic approach in Cariani, Kaufmann and Kaufmann (2013).

▶ *Standard motivation for information sensitivity:* Iris can bet on the outcome of a flip of a 90% biased coin at slightly worse than even odds (bet \$1, win \$1.9) or refrain from betting. The coin's bias is unknown. Iris only cares about money and values each dollar equally.

Judgments: We want (2)-(3a)-(3b) to be consistent:

(2) Iris should refrain from betting.

(3) a. If the coin is biased towards heads, Iris should bet on heads.
b. If the coin is biased towards tails, Iris should bet on tails.

A simple way of making them consistent: let the domain for *should* vary as a function of an underlying information state.

References

Available online: `cariani.org/SALT.pdf` (or ask over ☕).

4. Core Proposal

Thesis 1. Order alternatives (i.e. mutually incompatible sets of worlds), not worlds. (e.g.: {`refrain`, `bet-heads`, `bet-tails`} or {`drop`, `do not drop`}).

Thesis 2. Premises are pairs $\langle E, t\rangle$ with E a goal/desirable state and t a threshold. (slightly simplified compared to my paper, ☕)

Sample Premise: $\langle\{w \mid \text{Iris earns \$1 in w}\}, .8\rangle$

Intuition: an alternative A satisfies a premise $\langle E, t\rangle$ iff E (e.g. Iris earns \$1) is sufficiently probable (e.g., .8) given A.

Thesis 3. In ordering alternatives, the only part of A that affects its ranking is the part that overlaps the salient information/background.

Key definitions:

(i) a background state is a pair $\langle i, Pr\rangle$ consisting of a set of worlds i and a probability function Pr. An ordering source O is a set of premises (☕).

(ii) A satisfies $\langle E, t\rangle$ relative to $\langle i, Pr\rangle$ iff $Pr(E \mid A \cap i) > t$

(iii) $A \succeq B$ (relative to $\langle i, Pr\rangle$ and O) iff A satisfies (relative to $\langle i, Pr\rangle$) every premise in O that is satisfied by B.

Thesis 4. The domain of quantification for *should* is the set of w's in i that belong to the *maximal* alternatives (according to $\succeq$).

Thesis 5. Conditionals with non-probabilistic/non-modal antecedents restrict i.

5. Account of (2)-(3)

Assume context supplies appropriate premises [P1],[P2] and state $\langle i, Pr\rangle$. E.g.:

[P1] $\langle\{w \mid \text{Iris earns \$1 in } w\}, .8\rangle$; [P2] $\langle\{w \mid \text{Iris does not lose \$1 in } w\}, .7\rangle$
$Pr(\text{Iris earns \$1} \mid \text{Iris bets on heads} \cap i) = .5$
$Pr(\text{Iris earns \$1} \mid \text{Iris bets on heads} \cap \text{the coin is biased towards heads} \cap i) = .9$

Predictions: given that assignment, (2)-(3a)-(3b) are all predicted acceptable.

	in state i	in state ($i \cap$ "bias for heads")
`refrain`	[P2]	[P2]
`bet-heads`	-	[P1], [P2]
`bet-tails`	-	-
domain	`refrain` $\cap\, i$	`bet-heads` $\cap\, i \,\cap$ "bias for heads"

6. Account of (1a)-(1b)

Need a treatment for conditional antecedents of the form ⌜It is likely that φ⌝.

Option 1: To evaluate ⌜If it is likely that φ, ψ⌝ at $\langle i, Pr\rangle$ and O, evaluate ψ at $\langle i, Pr'\rangle$ and O where Pr' is a/the probability function that makes $[\![\varphi]\!]$ likely (for an approach roughly along these lines: Lassiter and Goodman, ms.).

Option 2: Context supplies a *set* of probability functions and updating on ⌜It is likely that φ⌝ rules out those that do not assign high (enough) probability to φ. (☕)

A dynamic system along the lines of Option 2 (based on Willer 2012, Yalcin 2012b and the semantics in §4) is on the 'details' handout or `cariani.org/SALT.pdf`.

Given either option, premises that work to derive (1a)-(1b) are:
[P1]$\langle\{w \mid \text{you pass the class in } w\}, .5\rangle$ [P2]$\langle\{w \mid \text{you avoid failing a class in } w\}, .5\rangle$

7. Why Not Scalar?

▶ The most important scalar theory draws on the notion of expected value.

For α a proposition, $EV(\alpha) = \sum_{w\in\alpha}[v(w) \cdot Pr(w \mid \alpha)]$

$[\![\textit{should } \varphi]\!]^{C,Pr,v,w} = T$ iff the expected value of $[\![\varphi]\!]$ (calculated on the basis of Pr and v) $>$ the expected value of the relevant alternatives.

[The above characterization is very rough: lots of implementations, each with advantages and disadvantages, ☕. See Goble (1996), Cariani (2009), Lassiter (2011, 2014). For a version on which *should* is a necessity operator: Wedgwood (2014)]

General Objection: the decision-theoretic notion of expectation should not be encoded by deontic semantics. This independence of semantics is reflected in specific predictions about attitudes of non-Bayesian agents.

Context: Suppose John believes that everyone ought to obey some odd decision rule, e.g. *Maximax* (i.e., choose the option that maximizes the maximum value).

(4) John thinks Iris should not refrain from betting. ✓

On a standard analysis for 'thinks', Expected Value theories do not derive this.

▶ **Operator analysis of 'thinks':** $[\![S \text{ thinks that } \varphi]\!]^{C,i,Pr,v,w} = T$ iff for every i', Pr', v' compatible with S's state, and all $w' \in I'$, $[\![\varphi]\!]^{C,i',Pr',v',w'} = T$.

What is v' that is compatible with John's state? Two possible options.

▶ **Option 1:** v' does not reflect John's risk-seeking disposition, e.g.:

non-reflecting v'	bias for heads (.5)	bias for tails (.5)	EV
`refrain`	0	0	0
`bet-heads`	1	-2	-0.5
`bet-tails`	-2	1	-0.5

Problematic Consequence: (4) is incorrectly predicted to have semantic value F.

▶ **Option 2:** v' does reflect John's risk-seeking disposition, e.g.:

reflecting v'	bias for heads (.5)	bias for tails (.5)	EV
`refrain`	0	0	0
`bet-heads`	2	-1	0.5
`bet-tails`	-1	2	0.5

Problematic consequence: (4) is predicted T, but we now are in trouble with:

(5) John thinks that if the coin is biased towards heads, Iris should bet on heads.✗

Intuitively ✗ because John thinks Iris must follow *Maximax* and does not care that the coin is biased (only cares about the maximum outcome).

But it is predicted T by the above scalar semantics (under this particular non-reflecting v': for my criticisms of other v' functions, ☕ or see the full paper on my website; for a response see Lassiter (2014)).

8. Account of (4)-(5)

If we implement the operator account in my system, there is no problem with attitudes of non-Bayesian agents.

▶ **Operator analysis of 'thinks':** $[\![S \text{ thinks that } \varphi]\!]^{C,I,Pr,O,w} = T$ iff for every i', Pr', O' compatible with S's state, and all $w' \in I'$, $[\![\varphi]\!]^{C,i',Pr',O',w'} = T$.

We may suppose that John's state is compatible with an ordering source containing only the premise: $\langle\{w \mid \text{Iris earns \$1 in } w\}, 0\rangle$. This prioritizes those alternatives that are compatible with Iris earning \$1.