# Two Ways to Want

Ethan Jerzak*
Draft, ask before citing

June 24, 2018

> No one, then, Meno, wants what is bad.
> –Socrates, *Meno*, 78a

**Abstract**

I present hitherto unexplored and unaccounted for uses of 'wants'. I call them **advisory** uses, on which information inaccessible to the desirer herself helps determine what it's true to say she wants. I show that extant theories by Stalnaker, Heim, and Levinson fail to predict it. I also show that they fail to predict true indicative conditionals with 'wants' in the consequent. These problems are related: intuitively valid reasoning with modus ponens on the basis of the conditionals in question results in unembedded advisory uses.

I consider two fixes, and end up endorsing a relativist semantics, according to which desire attributions express information-neutral propositions. The truth of a desire attribution depends on the state of information at the context of assessment. On this view, 'wants' functions as a precisification of 'ought', which exhibits similar unembedded and compositional behavior. I conclude by sketching a pragmatic account of the purpose of desire attributions, one that explains why it made sense for them to evolve in this way.

## 1 In vino veritas

Too often have I had the misfortune of being directed to bring wine to a dinner party. Not beer, or whisky, both drinks whose quality I'd be quite a bit more competent to judge, but specifically wine. With the aid of my visual system I can usually distinguish the red stuff from the white stuff; that just about exhausts my ability to make discriminations.

What I want is the best wine for the occasion, which I understand to be that which will bring the most joy to my more gustatorily advanced dining comrades. I only care about my comrades' taste; all wines taste fine to me. There I stand, in the grocery store, having whittled the options down to two. There's a Zinfandel from Sonoma Valley, and a Sauvignon Blanc from New Zealand. Unbeknownst to me, the Zinfandel would bring my dinner companions the most joy; they find the Sauvignon Blanc's grassiness oppressive. You, a maximally informed observer of the situation, are looking at me in my predicament. A natural way for you to describe the situation would be with (1):

*Thanks to Arc Kocurek, Hannes Leitgeb, John MacFarlane, Francois Reçanati, Rachel Rudolph, and Seth Yalcin for invaluable discussions and comments on drafts. Thanks also to audiences at UC–Berkeley, UCL, the University of Konstanz, the Institute Jean-Nicod, and the MCMP.

(1) He doesn't know it, but he wants the Zinfandel.

(2), however, rings false:

(2) #He doesn't know it, but he believes that the Zinfandel is the wine to get.

After all, if I believed that the Zinfandel were the wine to get, there would be no predicament—I'd simply get it, my comrades would savor it, and all would be well.

Suppose further that my dining companions change their taste, now finding the Sauvignon Blanc's grassiness pleasant and the Zinfandel's fruitiness overwhelming. (3) would then be the right thing to say:

(3) He doesn't know it, but he wants the Sauvignon Blanc.

Something strange has happened. Without changing anything about me—I've just been standing there dumbfounded all along—my desires seem to have changed. I went from wanting the Zinfandel to wanting the Sauvignon Blanc, without any corresponding change in my underlying psychological state.[1]

This contrasts with belief. Nothing about my beliefs has changed here. All along, I believe that whichever wine I buy should align with my comrade's preferences. What wine that *refers* to changes, but the content of my beliefs don't.[2] My beliefs about what's best to get are compatible with any situations in which the wine-to-get lines up with the wine-they-want. They mark no distinction between the Zinfandel and the Sauvignon Blanc. My desires, however, attributed in (1) and (3), seems to have a kind of sensitivity to the wider world that my beliefs do not. In situations where my comrades *actually* want the Zinfandel, whether I know it or not, there's a sense in which I want that too; and in situations where they actually want the Sauvignon Blanc, so, in some sense, do I.

[1] This kind of case was brought to my attention by Callard [2017], although she is concerned there not to give a particularly realistic semantics for natural language desire attributions, but rather to argue on behalf of Socrates that all we ever really want is the Good. This kind of use is also discussed in Davis [1984], and Rooryck [2017] discusses the differences in truth conditions between 'wants' and 'needs'. My interest here is in developing a formal account of how information states factor into the truth conditions of desire reports.

[2] Belief attributions sometimes bear *de re* readings, with behavior superficially similar to that of advisory desire attributions. If Susan has a general belief that all Minnesotans are nice, but no particular beliefs about some Minnesotan (Fred) whom she's never met or heard of, I could reasonably say, "Susan thinks Fred is nice". However, this phenomenon is more limited with 'believes' than with 'wants'. If Susan *had* met Fred, and, not knowing that he was Minnesotan, formed the definite opinion that there is nothing nice about Fred, such a *de re* belief report would be inappropriate. (1), on the other hand, is appropriate even if I'm completely (erroneously) convinced that my friends prefer the Sauvignon Blanc.

You might resist these data.[3] Perhaps (1) and (3) aren't really true. Perhaps all I ever really wanted all along mirrored my beliefs about what's best to get; the content of my desire was, less determinately, to get some wine or other that pleases my comrades. Someone pressing this line would insist that the correct way for you, the better-informed observer, to describe the situation would be:

(4) He doesn't want the Zinfandel. (Not yet anyway. But he will once he learns that it's the wine that his dinner companions prefer.)

I don't find (4) a horribly unnatural thing to say. But it's no more natural than (1). And (1) and (4) seem inconsistent. This suggests that we're attributing desires in two different ways. In (1), information beyond my ken helps determine what I want. In (4), what I want more or less coincides with what I believe to be good.[4]

(1) and (4) are typical examples of two different uses we make of 'wants'. One use is to predict and explain how agents will act, roughly along the lines of belief-desire folk psychology. If I know that someone wants $A$, and believes that doing $B$ will result in her getting $A$, and nothing stands in her way of doing $B$, I'll usually predict that she'll do $B$. The use of 'wants' in (1) clearly isn't this notion; if all you're allowed to do is observe, not to advise, you won't predict that I'll toddle off to the party with the Zinfandel in hand. Indeed, if you knew that I falsely believed my dinner companions to prefer the Sauvignon Blanc, you'd make exactly the opposite prediction. Let's call this the **predictive** use. On the predictive use, (1) is false and (4) is true.

But there's definitely another sense in which, if I buy the Sauvignon Blanc, I won't have bought what I really wanted all along. Indeed, I'd readily admit as much once my error becomes known to me. Say that I, falsely believing the Sauvignon Blanc to be preferred by my comrades, buy it and bring it to the party. It would be natural for me to express my regret with:

(5) Ach! That wasn't the wine that I wanted!

[3] Empirical work remains to be done to determine how cross-linguistically robust these advisory uses are. Informal surveys suggest that it is harder, if not impossible, to hear in (for example) German, Spanish, and French. An anonymous reviewer helpfully pointed out that the original meaning of 'want' in English simply meant 'lack', as in "the soup wants salt" or "you shall want for nothing". Only later did the psychological use develop. It could be this evolutionary history that explains why English is unique here, if it turns out to be. The interesting thing for our purposes is that there *is* an attitude verb in some language that exhibits the kind of information-sensitivity more commonly associated with modals.

[4] I say "more or less" in light of Lewis [1988]'s argument against such an identification. (Although see Bradley and Stefánsson [2016] for a counterargument.) The important point for these purposes, as will become clear in what follows, isn't the identification of desire with belief, but rather the identification of the *information* to which desire reports are sensitive with the desirer's own beliefs.

Or say that you, the maximally informed observer, break your silence to dispense advice. You'd say:

(6) Return that Sauvignon Blanc! That's not what you wanted, your comrades hate it. What you really wanted to buy was the Zinfandel.

It would be odd for me to retort:

(7) ?You're wrong! I really did want to buy the Sauvignon Blanc. I bought exactly what I wanted. But, now that I know that my comrades hate it, I *now*, having changed my mind, want the Zinfandel.

It would be much more natural to retract my previous claim about what I desired, saying something like:

(8) Oh! You're right, I guess I didn't want the Sauvignon Blanc after all. Thanks for telling me.

Situations like these, where better-informed agents offer advice to worse-informed ones, are where we most often find the use of 'wants' that I'm interested in. We ask the subway worker which train we want to get on, given where we're going; a good sommelier *tells* you what wine you want, instead of sitting back and laughing at you while you select the Chardonnay you erroneously thought would go nicely with your ribeye. This use of 'wants' isn't the predictive one. In telling you what you want, better-informed advisers like the subway-worker and the sommelier are making use of *their* information, not restricting themselves to yours. I'll call this the **advisory** use, since it figures most prominently in situations of advice.

In what follows, I take it as evident that we attribute desires in this advisory sense, and not just in fringe circumstances. Injunctions like, "Figure out what you really want, before you do anything you'll regret!" sound extremely natural, as do doubtful self-attributions, as in, "I think I want the 9am flight, but I won't know for sure until I know when the meeting is." Similar injunctions involving "believes" sound very weird. It's easier to be ignorant about what you want than about what you believe, and theories of attitude verbs shouldn't disallow that. My first task is to point out that the extant theories, being engineered with predictive uses in mind, don't predict advisory uses.

## 2 Warm-up: the naïve semantics

Here's a super flat-footed way to model desire attributions. Agents have, at bottom, preferences concerning outcomes, and what they want is a function of those preferences. To say that an agent wants $\varphi$ is just to say that the outcomes she most prefers are ones in which $\varphi$ holds.

A well-known problem for this approach, discussed in Stalnaker [1984], is that it predicts that I want whatever follows from, or is presupposed by, what I want. Say, for example, that John is sick, and would very much prefer not to be. It's true to say,

(9) John wants to get better.

But on the naïve semantics, this entails

(10) John wants now to be sick.

since every world in which John gets better is a world in which John is now sick. Therefore if "John gets better" is true throughout the worlds John considers best, "John is now sick" must also be true there. Thus if John wants to get better, he wants to be sick now. We'd expect him to protest this consequence, and our theory of desire attributions should not contradict him in this.

## 3 Stalnaker and Heim

This example shows that what we want isn't just a matter of what's going on in the worlds we most prefer. It also depends on which options are live in the situation we find ourselves in. John never wanted to be sick, but given that he is, he wants to get better. Thus what we want depends, in addition to basic preferences on outcomes, on a state of information—a state, that is, that includes certain options as live and rules out others as dead. It's our preferences regarding live options that factor into the truth conditions of a propositional desire report. Worlds in which John never got sick are not live options, so his preferences regarding them, strong though they might be, don't factor into characterizing his state of desire with respect to getting better.

How exactly does a state of information combine with basic preferences to yield desire attributions? A natural thought, first outlined by Stalnaker, is that I want $\varphi$ if, throughout the live worlds in the relevant state of information, my basic preferences render nearby $\varphi$ worlds better than nearby $\neg\varphi$ worlds. The question then becomes: which information is relevant? Hitherto the literature on desire attributions has implicitly assumed an answer to this question: The body of information that's relevant is that which characterizes the desirer's own beliefs. Stalnaker:

> Wanting something is preferring it to certain relevant alternatives, the relevant alternatives being those possibilities that the agent believes will be realized if he does not get what he wants. (Stalnaker [1984], 89)

Heim [1992], who fleshes out Stalnaker's idea formally, makes the same assumption. Some notation:

- $\succeq_x^w$: a preorder on worlds, so defined that $w_1 \succeq_{x,w} w_2$ just in case agent $x$ in $w$ weakly prefers $w_1$ to $w_2$ ($\succ_x^w$ for strong preference). For sets $W_1$, $W_2$ of worlds, $W_1 \succ_x^w W_2 := \forall w_1 \in W_1, \forall w_2 \in W_2, w_1 \succ_x^w w_2$;
- $B_x^w$: The set of worlds compatible with $x$'s beliefs in $w$;
- $\text{Min}_w(\varphi)$: The set of most similar worlds to $w$ in which $\varphi$ holds.

With these resources in hand, Heim proposes the following semantics:[5]

$$[\![x \text{ wants } \varphi]\!]^w = 1 \text{ iff } \forall w' \in B_x^w, \text{Min}_{w'}(\varphi) \succ_x^w \text{Min}_{w'}(\neg\varphi).$$

Neither Stalnaker's idea nor Heim's formalization of it was engineered with cases like (1) in mind. This is easy to see just by sketching a model faithful to the structure of our wine case and showing that Heim's semantics does not churn out (1). An explicit model of the case and a derivation of Heim's truth conditions relative to it are sketched in the appendix.

Intuitively, though, it's easy to see why Heim's semantics doesn't predict (1). In the *in vino veritas* case, I have no beliefs about which wine my comrades prefer. Thus, while my basic preferences render worlds where my selection aligns with my comrades' tastes better than those where it doesn't, my beliefs do nothing to single out the Zinfandel. So it's not the case that, throughout all worlds compatible with my beliefs, nearby I-buy-the-Zinfandel worlds are preferred by me to nearby I-buy-the-Sauvignon Blanc worlds. There are counterexamples among the non-actual worlds, which my beliefs do not rule out, where my comrades prefer the Sauvignon Blanc. Thus Heim's semantics misses true readings in situations like *in vino veritas*.

## 4 Decision-theoretic accounts

Levinson [2003] also complains that Heim's semantics fails to validate intuitively true desire attributions. But his cases are quite different in spirit from mine, and motivate a different kind of semantics from Heim's. Since I want a semantics that handles both kinds of cases (and, as we'll see in §6, combinations of them), it's instructive to consider his examples and the decision-theoretic semantics he cooks up to accommodate them. I'll then show that his semantics doesn't help with the *in vino veritas* case, and consider ways to improve on it.

[5] Heim actually casts her proposal in the framework of dynamic semantics; I've reformulated her view in a static setting, since the dynamic framework is motivated by considerations, orthogonal to the present ones, about the projection of presuppositions in attitude ascriptions.

Levinson's case against Stalnaker and Heim involves insurance. Most of us, he observes, want to buy insurance sometimes. Even though it's pretty unlikely that our houses will burn down, it would be such a calamity if they did, that many of us want to be safe rather than sorry. But this poses a problem for Heim. For consider two worlds where my house doesn't burn down, but which differ as to whether I bought insurance. On the whole, do I prefer the one where I bought insurance, or the one where I didn't? I, for one, prefer the world where I hold onto my cash, instead of shelling out for an as-it-happens useless insurance policy.[6] But loads of the worlds consistent with my beliefs are worlds where my house won't burn down irrespective of whether I buy insurance. Therefore, I don't meet Heim's requirement that all of my belief-worlds render nearby "I buy insurance" worlds better than "I don't buy insurance" worlds.

To figure out whether someone wants to buy a particular insurance plan, we need information more fine-grained than anything on offer in Heim's semantics. Full beliefs and qualitative preferences aren't enough; we need to know just how likely she judges it to be that her house will burn down, how bad it would be for her if it did, and what the plan costs. These are quantitative, not qualitative matters.

Thankfully we have a quantitative theory of rational action at our disposal: decision theory, which Levinson's semantics, following Goble [1996]'s account of deontic modals, is modeled after. Let's upgrade Heim's less fine-grained ingredients to the following more fine-grained ones:

- Upgrade $\succ_x^w$, a mere preorder on worlds, to an evaluation function $g_x^w : W \to \mathbb{R}$, defined such that $g_x^w(w_1) \geq g_x^w(w_2)$ just in case agent $x$ in $w$ (weakly) prefers $w_1$ to $w_2$.
- Upgrade the state of information, previously identified with the set of worlds $B_x^w$, to what Yalcin [2012c] calls a **sharp information state** $i_x^w = \langle S_x^w, Pr_x^w \rangle$:[7]
  - $S_x^w \subseteq W$ is the set of live epistemic possibilities for $x$ in $w$;
  - $Pr_x^w : \mathcal{A} \to \mathbb{R}_{[0,1]}$, for $\mathcal{A}$ a Boolean algebra of subsets of $W$, represents $x$'s credences over the live epistemic possibilities in $w$. $Pr_x^w(S_x^w) = 1$, and for disjoint $A, B \in \mathcal{A}$, $Pr_x^w(A \cup B) = Pr_x^w(A) + Pr_x^w(B)$.

[6] Büring [2003] defends Heim against Levinson by arguing that those who buy insurance *do* prefer worlds in which they buy unused plans, because as long as they don't *know* that the plan will be useless, they primitively value the peace of mind that insurance brings in such worlds. This is, of course, a formal possibility; but Büring then owes us a new substantive account of preference, and I have trouble seeing how it could account for all insurance-style cases. There are gamblers and actuaries who make claim to make bets dispassionately, in the sense that they are perfectly psychologically at ease gambling and losing so long as the gamble was rational given their utilities and credences. That is, they explicitly claim not to primitively value peace of mind. It's hard to see how Büring could account for such cases, whatever substantive account of preference he gives. See Lassiter [2011] for further arguments in favor of a more fine-grained probabilistic framework.

[7] I stipulate that the set $W$ of worlds is finite. Otherwise we would have to switch from sums to integrals, or from probability functions defined on worlds to those defined on partitions of the set of worlds. Since all of the cases we'll be interested in involve a finite number of worlds/outcomes, these extra complications are unnecessary.

- Shorthand: $Pr_x^w(w' \mid [\varphi])$: $x$'s credence in $w$ that $w'$ is the actual world conditional on $[\varphi] = \{w : [\![\varphi]\!]^w = 1\}$.

Levinson proposes a semantics which says that you want $\varphi$ just in case, relative to your credences and utilities, $\varphi$ yields higher expected utility than $\neg\varphi$.[8] Formally,

$$\begin{aligned}[\![x \text{ wants } \varphi]\!]^w = 1 \quad &\text{iff} \quad EU_{x,w}(\varphi) > EU_{x,w}(\neg\varphi) \\ &\text{iff} \quad \sum_{w' \in S_x^w} g_x^w(w') Pr_x^w(w' \mid [\varphi]) \\ &\qquad > \sum_{w' \in S_x^w} g_x^w(w') Pr_x^w(w' \mid [\neg\varphi]).\end{aligned}$$

Levinson [2003] sketches an explicit model of the insurance case, and shows how his semantics predicts that 'you want to buy insurance' is true relative to it. Intuitively, while my full beliefs don't rule out that I'm in a situation where I shell out money for an as-it-happens useless plan, my quantitative preferences render an uninsured fire-destroyed house to be so calamitous an eventuality that, even though I judge it to be pretty unlikely, the calamitousness overwhelms the slim odds, making it worth shelling out a relatively small amount of money.

Thus Levinson predicts what Heim fails to predict—that I can want $p$ even if not *all* of my belief worlds are ones where I prefer nearby $p$ worlds to nearby non-$p$ worlds. This is a virtue of his account. Plus, the decision-theoretic framework also easily generalizes to graded desire attributions ("I *really* want beer"; "I want beer, but I want whisky way more"), whereas it's hard to see how Heim would have the tools for this.[9]

Does Levinson's semantics help with *in vino veritas*? Again, a quantitative model and derivation of truth conditions relative to a true-to-case model is sketched in the appendix. However, it's again easy to see intuitively that Levinson's semantics won't help with *in vino veritas*. Just as my full beliefs and qualitative preferences don't change depending on my interlocutors' information, neither do my credences and quantitative preferences. Relative to my credences and utilities, I expect to be no better off buying the Zinfandel than buying the Sauvignon Blanc. Indeed, even if my credences and utilities rendered buying the Sauvignon Blanc the preferred action, the store adviser, having better information, can still felicitously correct my desire report. After he does this, I should retract any assertions to the effect that I wanted the Sauvignon Blanc. So Levinson, while improving on one aspect of Heim's proposal, does not solve our problem about advisory desire reports.[10]

[8] This is a slight simplification of Levinson's official view. He actually defines 'wants' relative to evaluation functions $g$, in order to handle cases of active ambivalence between outcomes resulting in seemingly contradictory desire attributions. (E.g. "I want the wine [it will taste great], but I also don't want it [it will cause a hangover].") This is a different problem from the kind I'm interested in—in the *in vino veritas* case, what's going on clearly isn't that you change how you feel about total outcomes, but rather that, given your fixed total preferences, different information states yield different results about what you want.

[9] See Lassiter [2011] for a probabilistic account of modality that incorporates scales, familiar from the literature on gradable adjectives, to account for these data. See also footnote 13 for some problems with this approach.

[10] Other proposals for the semantics of 'wants' exist: for example, those of von Fintel [1999], van Rooij [1999], Villalta

## 5 'Wants' in the consequent of conditionals

The above shows that some true sentences involving 'wants' in certain contexts come out false on the theories offered by Heim and Levinson. In this section I'll show that their theories also do not predict certain aspects of its compositional behavior. Back to the wine. Consider the following conditionals, said by you of me in the *in vino veritas* case:

(11) If his comrades prefer the Zinfandel, he wants the Zinfandel.

(12) If his comrades prefer the Sauvignon Blanc, he wants the Sauvignon Blanc.

These are both not only true, they are *extremely* true, in that they're among the very most natural ways to describe the state of mind I'm in when I'm standing there dumbfounded in the store. Note here again the contrast with belief. (13) is extremely false:

(13) If his comrades prefer the Zinfandel, he believes that the Zinfandel is the wine to get.

You can use (11) and (12) to describe my conditional preferences, but (13) cannot be used to describe my conditional beliefs. (13) means that my beliefs are sensitive to my comrades' preferences, which, as a feature of the *in vino veritas* case, they are not. Granted, if I use a version of (13) first-personally, it doesn't sound *too* bad: "If my comrades prefer the Zinfandel, I think that's the wine to get."

But third personally it clearly doesn't work. To see this, consider a more knowledgable third party engaging in a bit of reasoning about what you want/believe. He would do ill to reason:

> If his comrades prefer the Zinfandel, he believes that the Zinfandel is the wine to get.
> His comrades prefer the Zinfandel.
> ---
> He believes that the Zinfandel is the wine to get.

My comrades do prefer the Zinfandel, but I don't believe that the Zinfandel is the wine to get. The most plausible diagnosis of why this is bad reasoning is that the major premise is false; my beliefs aren't sensitive to my comrades' preferences, as it requires. However, given the availability of the advisory 'wants', he would do well to reason:

> If his comrades prefer the Zinfandel, he wants the Zinfandel.
> His comrades prefer the Zinfandel.
> ---
> He wants the Zinfandel.

[2000], Lassiter [2011], and Condoravdi and Lauer [2016]. The differ in details, but all of them are fundamentally engineered to take the subject's doxastic state as the information relative to which the desire attribution is assessed.

Indeed, this is exactly the kind of reasoning you'd engage in if wondering which bottle you should hand me.

This suggests that the maybe-vaguely-true-ish first-personal version of (13) is interpreted with the "thinks" taking wide scope over the conditional. This response is not available for (11) and (12), however. It would have it that sentences superficially of the form

$$\varphi \rightarrow x \text{ wants } \psi$$

are to be interpreted as

$$x \text{ wants } (\varphi \rightarrow \psi).$$

This approach has several shortcomings, of which I'll mention two. First, it doesn't validate the intuitively valid reasoning above, which results in your concluding that I want (in the advisory sense) the Zinfandel, as an instance of *modus ponens*. Perhaps the semantics of 'wants' could be fiddled with in such a way as to make $\{p, x \text{ wants } (p \rightarrow q)\}$ entail $\ulcorner x \text{ wants } q \urcorner$, but this also wouldn't be valid on Heim's or Levinson's semantics without modification. Since we'll need to modify the semantics anyway to make sense of the truth of these conditionals and the ability to reason with them using *modus ponens*, we might as well not butcher the surface grammar.

Second, the strategy crashes when the consequents are truth-functionally complex. Consider:

(14) If his comrades prefer the Zinfandel, then he wants to buy the Zinfandel, and (/but) they are snobs.

It's not clear how a defender of wide-scoping could interpret mixed conditionals like this. You might try:

$$me \text{ wants } (p_z \rightarrow (b_z \wedge snobs))$$

But this would be false—not wanting snobs for friends, but taking it to be quite possible that they prefer the Zinfandel, I certainly don't want it to be the case that, if my friends prefer the Zinfandel, they are snobs.[11] The best and simplest explanation here is that (11) and (12) are true, and have the logical form they seem to have.

Here's why Heim's and Levinson's accounts do not yield (11) and (12). I'll give a working semantics for the indicative conditional and show that (11) and (12) don't come out true in a

[11] Something like this argument is present in Kolodny and MacFarlane [2010] for conditionals involving 'ought', and it traces back to Thomason [1981]. The same mixed conditional would tell against an attempt to treat 'wants' as a primitive dyadic operator, of the form $\ulcorner x \text{ wants } (\varphi \mid \psi) \urcorner$. In general, the dialectic here mirrors the dialectic involving the interaction between deontic modals and conditionals. This, I argue in §8, is no accident, but illustrates deep structural similarities between 'wants' and 'ought'.

moment. But first an informal gloss: A conditional is true in a context when, suppositionally adding the antecedent to the stock of information at that context, the consequent comes out true under that hypothesis. So add to the common information in a case like *in vino veritas* that my comrades prefer the Zinfandel. Is it true that I want the Zinfandel, on the semantics given by Heim or Levinson? No—adding that information doesn't instruct us to change anything about my credences/beliefs or preferences/utilities. What I believe and prefer just depends on the world, not on the state of information in the common ground. So roughly speaking, the consequent will have the same truth conditions in the updated information state as in the non-updated one, and we've already seen that it's false with respect to those truth conditions in cases like *in vino veritas*.

Formally, I'll adopt a working semantics for $\rightarrow$ as a kind of epistemic modal. On this view, developed and defended by Yalcin [2007], Kolodny and MacFarlane [2010], and MacFarlane [2014], a conditional functions as a test on the stock of information mutually presupposed in the conversational context: it tests whether adding the antecedent to that stock ensures that the consequent is true. Indicative conditionals are assessed relative to worlds and bodies of information $i$. Some definitions will be helpful. Shorthand: $[\varphi]_i = \{w \mid [\![\varphi]\!]^{w,i} = 1\}$.

**Definition.** An information state $i$ **accepts** $\varphi$ iff $[\varphi]_i = S_i$. In other words, iff $\forall w \in S_i,\ [\![\varphi]\!]^{w,i} = 1$.

**Definition.** The information state $i$ **updated by** $\varphi$, written $i + \varphi$, is $\langle S_i \cap [\varphi]_i, Pr_i^{\varphi} \rangle$, where $Pr_i^{\varphi}(x) = Pr_i(x \mid [\varphi]_i)$.

The semantics for the indicative conditional $\rightarrow$ is then:

$$[\![\varphi \rightarrow \psi]\!]^{w,i} = 1 \quad \text{iff} \quad i + \varphi \text{ accepts } \psi.$$

It's a straightforward matter to verify (see appendix) that (11) and (12) both come out false on Levinson's semantics, relative to realistic models of *in vino veritas*.

What kind of account might help predict such uses? When we use conditionals like (11) and (12), we're describing something like my conditional preferences. Roughly speaking, (12) describes my state of mind when, restricting my attention to worlds in which my comrades prefer the Sauvignon Blanc, my preferences and *updated* credences judge I-buy-the-Sauvignon Blanc worlds to be better. To predict these truth conditions, we'll need the semantic value for 'wants' to be sensitive to the state of information that indicative conditionals operate on. That way, the antecedents of conditionals can modify the information parameter in the semantic entry for 'wants' in the right way. This suggests that 'wants' belongs to the class of informational modals like epistemic might/must, deontic ought/may, and probability operators.[12] Indeed, I argue in §8, it functions as a systematic precisification of 'ought'.

[12] Although see von Fintel [2012] for a defense of more classical approaches to these phenomena.

I'll sketch two different proposals. The first posits a lexical ambiguity in 'wants': a predictive entry governed by a semantics like Levinson's, and a "perfect information" entry which relativizes the information parameter to the state of perfect information at a world. I'll sketch some reasons for dissatisfaction with this bifurcation response, and then propose the semantics I'll ultimately endorse, according to which desire attributions express information neutral propositions.

## 6 Overreaction: perfect information

A natural reaction here would be twofold. First, since 'wants' does seem to have a sense, namely the predictive sense, more or less consonant with Levinson's semantics, one might posit a lexical ambiguity and use Levinson's semantics for 'wants$_{pred}$'. Second, one would add a new semantic entry for the advisory sense, 'wants$_{advise}$'. This semantics would have it that we want$_{advise}$ whatever our preferences judge to be better, not according to the state of information which characterizes our incomplete and possibly defective beliefs, but rather according to the state of perfect information. We really want what will *actually* put us into preferred worlds, in light of all facts known and unknown. That would suggest something like the following:

$$[\![x \text{ wants}_{advise} \varphi]\!]^w = 1 \text{ iff } Min_w(\varphi) \succ_x^w Min_w(\neg\varphi).$$

This semantics can predict the data of *in vino veritas*: relative to the actual world, nearby "I buy the Zinfandel" worlds are better according to me than nearby "I buy the Sauvignon Blanc" worlds. It can also predict the conditionals we've been interested in. Start out with a state of information that doesn't settle which wine my comrades prefer, and then update it with "my comrades prefer the Sauvignon Blanc." Relative to the worlds in this updated state, nearby worlds in which I buy the Sauvignon Blanc are better than those in which I buy the Zinfandel. So as far as the considerations on the table so far go are concerned, this bifurcation proposal has everything going for it.

However, this response is an overreaction that we should reject for two reasons. First, it would make the advisory sense extremely difficult to justifiably use. Second, it can't account for true advisory uses in situations of known uncertainty—essentially, when Levinson-style insurance cases involve advisory aspects due to disagreement about the likelihoods of the relevant outcomes.

The first problem is simply that perfect information isn't easy to come by. To confidently assert that I want $\varphi$ in the advisory sense, the PI semantics has it that you have to be fairly confident that, taking absolutely every consequence of your action throughout all time into account, I'll be better off by my own lights if $\varphi$ than if $\neg\varphi$. That's quite a claim. Sure, you know that my comrades prefer the Zinfandel. But maybe but they are in such good spirits today that if I buy the Zinfandel,

the party will be too rambunctious and we will all miss work tomorrow. Then I'd want$_{advise}$ *not* to buy the Zinfandel. But maybe in addition to this all of our bosses will have taken the day off, and missing the day will have no immediate consequences. Then I'd want$_{advise}$ the Zinfandel after all. But maybe, in addition to all of this, missing one day without consequence will instill in us a cavalier attitude towards punctuality, causing problems in our personal and professional lives. In this case, I don't want$_{advise}$ the Zinfandel. And so on.

It might be claimed that this isn't so bad, since usually I can be reasonably confident, if never totally certain, that only relatively normal consequences will follow from my comrades' enjoying a nice bottle of wine. So maybe we can never know for sure the truth of an advisory desire attribution, but we can often be justified in asserting them, and they can often turn out true.

But (the second problem) this simply gets the wrong result when I responsibly use the advisory sense in cases where I don't have perfect information, and I'm perfectly aware that my advice probably conflicts with what those with perfect information would advise. Take a modified version of a Levinston-style insurance case:

> **Insurance-Arsonists:** You just declined to buy an insurance plan, because according to your credences and utilities, it was just barely too expensive to be worth it. However, I, unlike you, happen to know that a gang of arsonists has just moved to town. Thus the probability of your house burning down is much higher than you think it is—enough to tip the scales back in favor of your buying the plan. You've just finished telling the insurance salesman that you don't want the plan.

I speak truly when I say to you (in a whisper, naturally, so as not to tip off the lingering insurance salesman that his plan is probably mispriced):

(15) "No, that's wrong—you actually do want to buy this plan. I'll explain why later."

Now, as it happens, the world is such that the gang of arsonists will spare your house. So your house won't burn down, and you lose the money you spent on the plan. And, remember, you prefer no-housefire worlds in which you didn't shell out for the plan to ones where you did. Thus if wants$_{advise}$ is relativised to perfect information, I speak falsely in (15). This seems wrong. (15) seems like true and excellent advice, at least when I make it.

It's open to maintain that (15) *is* false, but to explain its seeming like good advice by holding that I was justified in asserting it. But it's hard to see why I would be justified in asserting it, if 'wants$_{advise}$' has these truth conditions. After all, when I assert (15), I know that it's still more probable than not that your house won't burn down, marauding arsonists notwithstanding. The arsonists aren't *that* efficient. Thus if the semantics of 'wants' in (15) were given by perfect

information, I should think that (15) is very probably false when I assert it. So it's very difficult to see how I could nonetheless be justified in doing so.

The Insurance-Arsonists case suggests two things—one about the *source* of the information states that factor into the semantic values of advisory desire reports, the other about their *structure*. First, it suggests that the source of these information states isn't something that we can simply read off of the world of utterance. When advising people about what they really want, we aren't committing ourselves to something that only omniscient beings could know—that, taking account of absolutely every downstream consequence, you'll prefer the worlds that will/would result if the ascribed desire comes/came out true, compared to those in which it comes/came out false. The source of this information is more limited, and plausibly depends on context in some way.

Second, this case suggests that, whatever the *source* of these information states, their structure must be more fine-grained than that of Heim's semantics: they must represent some notion of likelihood, combined with a more fine-grained representation of preference. In the Insurance-Arsonists case, the metaphysically most similar worlds to ours in which you buy insurance are still worlds where your house does not burn. This is so even relative to the worlds doxastically accessible to me, the advisor. My information differs from yours not in terms of brute doxastic possibilities vis-a-vis house-burning: *both* of our doxastic possibilities include some housefire worlds and some no-housefire worlds, regardless of whether insurance is bought. In *neither* case will a semantics based on Heim's predict, even relative to the advisor's information, that you want to buy insurance. But this is wrong; my probabilistic information *can* make a difference to the truth value of a desire report. Thus whatever more flexible information base we relativize desire attributions to, that information base must include some representation of likelihood.[13]

One final reason to think that probabilistic structure is unavoidable, even on a more flexible account of the information source: desire ascriptions interact in non-trivial ways with probability operators in the antecedents of conditionals. Say that your roommate Ahmed, caring about your well-being and contemplating the possibility of rain, is advising you about whether to take an umbrella. There are two umbrellas in the house: a large and very effective one, and a small and moderately effective one. Your roommate is concerned about your not getting wet, but also about

[13]See Lassiter [2011] for further motivations for decision-theoretic semantics for a variety of modals. I do have some reservations about the standard EU approach here. For one thing it builds a huge amount of probabilistic and preferential coherence into the very meaning of desire reports, in a way that seems implausible; see Buchak [2013] for discussion. What I take Insurance-Arsonists to show is that information states, even for advisory uses, must include some representation of likelihood. I've chosen the EU framework of Levinson [2003] because it's by far the most well-known account. There are less committal alternatives: see Holliday and Icard [2013] and Holliday, Icard, and Harrison-Trainor [2017]. It's plausible that a more permissive theory would be more realistic, but delving into that more complicated machinery would unnecessarily cloud matters here.

your not traipsing around unnecessary weight. We might communicate his desires concerning which umbrella you should take as follows:

(16) If it's probably not going to rain, Ahmed wants you not to take any umbrella.

(17) If it's probably going to rain, Ahmed wants you to take the small umbrella.

(18) If it's going to rain, Ahmed wants you to take the big umbrella.

Conditionals like these are easy to account for if the information states relative to which advisory desires are assessed have probabilistic structure. On a framework like Heim's, it's hard to see how such an account would go, since she only has qualitative doxastic possibilities in her toolbox.

## 7 Information-neutral desires

What, then, is the source of the information states that factor into the semantics of desire reports? It is not necessarily the desirer's: advisers can help themselves to information beyond that of the attributee herself. But it is not, as Socrates plausibly claimed, the omniscient information state. The information states that license even advisory desire attributions should still be human-sized, so to speak, and sensitive to probabilities and utilities in the way suggested by the Insurance-Arsonists case.

One could develop a contextualist semantics that indexes the information state to the attribut*er*'s information, but this is unpromising, for it wouldn't explain the genuine *disagreement* we seem to be in when we disagree about what someone really wants. If the proposition I express when *I* use the advisory 'wants' is indexed specifically to my information, and yours is specifically indexed to yours, then we simply talk past each other when we disagree. On the contextualist account, if you and a third party disagreed about which wine my comrades preferred, we should be happy to have the following exchange:

> You: "He wants the Zinfandel."
> Third party: "Well, yes, I agree, but he doesn't want the Zinfandel."

That should sound just as good as a long distance phone conversation running:

> You: "It's raining here."
> Third party: "Well, yes, I agree, but it's not raining here."

But it doesn't sound just as good. We're not talking past each other; we have genuinely incompatible views about what the agent really wants, not compatible views about what would put the agent in preferred states according to our respective information.

This dialectic is quite reminiscent of extant debates on epistemic and deontic modals. The most promising options for such information-sensitive vocabulary are some sort of flexible/group contextualism (Dowell [2011] and [2013]), expressivism (Yalcin [2012b]), and relativism (Kolodny and MacFarlane [2010]). For the sake of predictive concreteness, I'll sketch a relativistic version here, but my semantics can be easily adapted to expressivist or contextualist background theories.

My proposal has two features. First, I'll model probabilistic informational common grounds with blunt probabilistic information states. Second, I introduce what I call 'mixed' expected utility functions $EU^i_{g^w_x}$, where the utilities come from one source (the agent $x$ in world $w$), and the probabilities come from another (the blunt information states $I$ representing the probabilistic common ground of the conversation). I'll explain these elements in turn.

First, the blunt information states relative to which semantic values of formulas are assigned are:

**Definition.** A **blunt information state** $I$ is a set of sharp information states $i = \langle S_i, Pr_i \rangle$, such that they agree on all the coarse-grained possibilities: $\forall i_1, i_2 \in I,\ S_1 = S_2$. (So it makes sense to speak of $S_I$.)

My proposal says that you want what yields highest expected utility according to your utilities, combined not with your credences, but instead with the probabilities of the information state in the common ground. First, a definition:

**Definition.** The **mixed expected utility** of $\varphi$, $EU^i_g(\varphi)$, relative to a utility function $g$ and sharp information state $i = \langle S_i, Pr_i \rangle$, is the expected utility of $\varphi$ derived from the probability function of $i$ and the utility function $g$:

$$EU^i_g(\varphi) := \sum_{w' \in S_i} g(w') Pr_i(w' \mid [\varphi]_i)$$

My semantics uses these mixed functions. It goes:

$$[\![x \text{ wants } \varphi]\!]^{w,I} = 1 \quad \text{iff} \quad \forall i \in I, EU^i_{g^w_x}(\varphi) > EU^i_{g^w_x}(\neg\varphi).$$

I will ultimately endorse this semantic entry, together with a relativistic postsemantics running as follows (see MacFarlane [2014] for a general explanation of this relativistic framework):

> An utterance of the form $\ulcorner x \text{ wants } \varphi \urcorner$ is true as used at $c_1$ and assessed from $c_2$ if and only if $[\![x \text{ wants } \varphi]\!]^{w_{c_1}, I_{c_2}} = 1$.

This relativistic package, I'll argue, can predict the problematic data, and isn't saddled with the undesirable baggage of the bifurcation, perfect information response. I won't explain the entire relativistic semantic apparatus from the ground up—for that, see Egan [2007], Bledin [2014], and

MacFarlane [2014]. Instead, I'll walk through the kind of predictions this package makes in this case. These predictions, I'll argue, are supported by the data, providing confirmation for this kind of approach.

### 7.1 Relativistic *veritas in vino*

According to this theory, desire attributions require two contexts to be assessed true or false: the context of use, and the context of assessment. So to judge the theory, we have to give a bit more information about who is asserting (1), and who is assessing it, in what kind of context.

Say that, in a context where it's erroneously taken for granted that my comrades prefer the Sauvignon Blanc, I say to myself:

(19) I want the Sauvignon Blanc.

This is true as used and assessed relative to $c_1$, the context in which I utter it. This explains why I am justified in doing so. Now suppose that, later on in the shopping trip, you, having overheard (19), say:

(20) What you said before [in 19] is actually wrong—you don't want the Sauvignon Blanc, you want the Zinfandel. That's the one your comrades prefer.

The relativistic semantics judges that, in this new context $c_2$, you are right; you've changed the context to include the information that my comrades prefer the Zinfandel. That means that (19), as used at $c_1$ and assessed at $c_2$, is false; relative to this better information, I want the Zinfandel, not the Sauvignon Blanc. Thus this package predicts that I'm obligated to retract (19), once I learn that my comrades prefer the Zinfandel. This is the correct result; the data of §1 illustrates that it sounds very weird for me to stand by assertions like (19), once I acquire information relative to which my preferences render the opposite result. But it also predicts why it made sense for me to assert (19); assessed relative to the context of assertion, what I said was true.

It also predicts, as the perfect-information semantics did not, the right results in the modified insurance case. When I learn about the marauding arsonists, my credence that nearby houses will burn rises. So when I whisper to you that you're wrong about wanting to decline the plan, I speak truly, relative to your context of utterance and my context of assessment. While *your* credences render the plan too expensive to be worth it, your utilities mixed with *my*, the assessor's, credences render the plan worth the money after all. I'm not asserting, falsely, that you *will* be better off buying the plan. I'm saying that it's the best option, relative to your utilities and what I know to

be better information. That's why I give true and excellent advice when I tell you that you really want to buy the plan. Of course, if an even better-informed third party came along who knew that the arsonists planned to spare your house, then I should retract my assertion to the effect that you want the plan, and you should retract your retraction. This all jives extremely well with the information-neutral semantics, and wouldn't be predicted on the perfect information view.

## 7.2 Conditionals

Let's look at how the relativistic framework deals with the conditionals that were problematic for Heim and Levinson. Remember the conditionals:

(21) If his comrades prefer the Zinfandel, he wants the Zinfandel.

(22) If his comrades prefer the Sauvignon Blanc, he wants the Sauvignon Blanc.

The semantics for the indicative conditional carries over exactly from before, modified in a super-valuationist spirit to accommodate blunt probabilistic information states:

**Definition.** A blunt information state $I$ accepts $\varphi$ iff $\forall i \in I$, $i$ accepts $\varphi$.

**Definition.** The blunt information state $I$ **updated by** $\varphi$, written $I + \varphi$, is $\{\langle S_i \cap [\varphi]_I, Pr_i^{\varphi}\rangle \mid i \in I\}$, where $Pr_i^{\varphi}(x) = Pr_i(x \mid [\varphi]_I)$.

The semantics for the indicative conditional $\rightarrow$ is basically unchanged:

$$[\![\varphi \rightarrow \psi]\!]^{w,I} = 1 \quad \text{iff} \quad I + \varphi \text{ accepts } \psi.$$

The kind of information states where (11) and (12) are paradigmatically asserted are ones which include open worlds where my comrades prefer the Zinfandel, and open worlds where my comrades prefer the Sauvignon Blanc. I provide in the appendix a particular such information state, and show that the conditionals come out true. But again, intuitively, it's not hard to see what's going on. The antecedent of an indicative conditional like (11) restricts our attention to worlds in which my comrades prefer the Zinfandel, and asks what my expected utilities are, relative to information states including only *those* worlds, between my buying the Zinfandel and my buying the Sauvignon Blanc. Relative to these information states, my utilities render buying the Zinfandel the better option. So the indicative conditional is true relative to the original information state. *Mutatis mutandis* for (12).[14]

[14]See the end of the appendix for an consequence relation tracking information preservation, drawing on Yalcin [2012a], Willer [2012], and Bledin [2014], on which modus ponens comes out valid. Interestingly modus tollens fails, and this is a good thing: In a context where we're ignorant about which wine my comrades desire, the following can plausibly all be truly asserted: A. If my comrades prefer the Zinfandel, I want the Zinfandel. B. It's not the case that I want the Zinfandel. C. My comrades prefer the Zinfandel. The situation is similar to that of Yalcin [2012a]: the

So, this relativistic semantics can predict the assertability/retraction data of the *in vino veritas* case. We've also seen that it, together with a plausible semantics for the indicative conditional, can predict conditionals like (11) and (12) in the contexts in which they seem true. Thus this relativistic theory has two predictive marks in its favor over previous semantics, without falling prey to the inadequacies of the perfect information, bifurcation response.

## 8 'Wants' and 'Ought'

On the view I've offered, desire attributions function not only to predict what agents will do, but also to advise them about what courses of action they should undertake, if they want to realize their aims. To assert that someone wants $\varphi$ is to claim that, relative to her preferences and the best information available, she'll be better off by their own lights bringing about $\varphi$ rather than $\neg\varphi$. That's not far from what we sometimes communicate with 'ought'. Telling an agent what she really wants is basically a way of telling her what she ought to do, given her basic aims, but not necessarily limited to her information about how to achieve those aims.

This similarity is unsurprising, for 'wants' and 'ought' exhibit similar puzzling behavior. Just a few examples:

- Ross' puzzle:
  - $x$ wants $\varphi \not\Vdash x$ wants $(\varphi \vee \psi)$;[15]
  - $x$ ought to $\varphi \not\Vdash x$ ought to $(\varphi \vee \psi)$.
- Puzzling assertion/retraction data:
  - Both $\ulcorner x$ wants $\varphi\urcorner$ and $\ulcorner x$ ought to $\varphi\urcorner$ sound fine to assert if, relative to the common information at the context of assertion, $x$ can expect to be better off by her own lights supposing $\varphi$ than supposing $\neg\varphi$; but such assertions must be retracted if new information comes to light under which the opposite holds.
- Puzzling interaction with conditionals:
  - Both $\ulcorner\varphi \rightarrow x$ wants $\psi\urcorner$ and $\ulcorner\varphi \rightarrow x$ ought to $\psi\urcorner$ can be used to express conditional obligations/desires, motivating views of the indicative conditional as a kind of modal restrictor. (Kolodny and MacFarlane [2010], Yalcin [2012a], Bledin [2014])

On my view, 'wants' is a precisification of 'ought', which clarifies the kind of advice that is being given to agents. 'Ought' has notoriously many senses. If I claim that you ought to $\varphi$, I could

conditional is true in virtue of what *would* happen to the state of information after updating by the antecedent of the conditional; the desire attribution is false because relative to the original, more ignorant state of information, the Zinfandel and the Sauvignon Blanc yield equal expected utility; and the statement of my comrades' actual preferences is just a plain fact about the world.

[15] Ross' puzzle is solved on my decision-theoretic semantics. A state of information and utility function could give $\varphi$ higher expected utility than $\neg\varphi$, while failing to give higher expected utility to $\varphi \vee \psi$ than $\neg(\varphi \vee \psi)$.

be trying to communicate one of at least three things. I could be communicating that the better thing for you to bring about, given your subjective preferences and your subjective information, is $\varphi$ rather than $\neg\varphi$. ("Oh well—even though the gamble didn't pay off, you did what you ought to have done.") Or I could be communicating that, relative to your preferences, but *my* information, your basic ends are more likely to be achieved by bringing about $\varphi$ rather than bringing about $\neg\varphi$. ("Stop, you ought not buy the Sauvignon Blanc! Even though I hate it and think that everyone who prefers it is a snob, your comrades will be much happier with the Zinfandel, and that's what you care about.") Or I could be communicating my disagreement with your ends themselves, represented by your preferences on worlds. ("You ought to buy the Sauvignon Blanc, even though your comrades hate it! Your comrades are snobs.") My theory of desire attributions predicts that only the first two of these three meanings is available for 'wants'.[16]

This prediction is supported by data about how 'wants' and 'ought' embed differently under other attitude verbs. I'll focus here on 'thinks'. Consider Fred, a fellow dining comrade in the *in vino veritas* case. Fred knows that my comrades prefer the Zinfandel. He alone prefers the Sauvignon Blanc, and furthermore he is a solipsistic hedonist; he thinks that only his preferences should be taken into account when people are deciding what to do. My basic preference is to please as many of my comrades as possible, without any special provision for Fred. What does Fred think about all of this? I could describe Fred's attitudes as follows:

(23) Fred thinks that, although *I* think I want to buy the Sauvignon Blanc, I actually want to buy the Zinfandel.

After all, he knows my preference is to please the majority of my comrades, and he knows that my comrades prefer it. However, it doesn't seem right to say:

(24) Fred thinks that, although *I* think I ought to buy the Sauvignon Blanc, I actually ought to buy the Zinfandel.

If Fred thinks that I ought to buy the Zinfandel, then Fred *himself* prefers that I buy the Zinfandel. But Fred doesn't prefer this; he's a solipsistic hedonist, and only cares about getting his treasured Sauvignon Blanc. He thinks I *ought* to buy the Sauvignon Blanc, even though what I *really* want

[16] Schroeder [2011] also contrasts 'wants' with 'ought', but the differences he highlights are orthogonal to those that I'm interested in. He points out that 'wants' functions as a control verb, while 'ought' is ambiguous between a control verb (which builds in an agent, as in "John ought to ski") and a raising verb (which operates solely on propositions, as in, "it ought to be the case that John skis"). The differences I highlight arise as a distinction between 'wants' and the control verb sense of 'ought'.

is to buy the Zinfandel, even though what I *think* I want is to buy the Sauvignon Blanc.

That suggests that the point of having an advisory 'wants' is to have a linguistic device that behaves like 'ought' with respect to information, but which rigidly fixes the agent whose preferences we're evaluating the relevant possibilities with respect to. A claim using 'ought' leaves undetermined whether I'm adding to the common ground my own information, or my own preferences, or both; a claim involving the advisory 'wants' clarifies that I'm only concerned with the information component. Thus the advisory 'wants' clarifies the *kind* of advice I'm giving the agent. Whereas 'ought' can give moral advice about how the agents' preferences should ideally go, 'wants' can only give pragmatic advice about how agents can best achieve their given aims.

## 9 Whither the predictive 'wants'?

I haven't said much about the predictive sense of 'wants', the only sense hitherto accounted for in the literature. What's the relation between predictive uses and advisory uses?

The first thing to point out is that my modification to Levinson's semantics is, in many ways, very conservative. Usually—not always, but usually—agents are aware of what consequences various actions are likely to bring about. In a large number of central cases, the attributee of a desire attribution is in more or less the same state of information as the attributers. Thus predictive and advisory uses can be expected to coincide in tons of cases. I attribute to you a desire to have one of the beers in the fridge; rarely do I have unique access to evidence that the beer is poisoned, or that the refrigerator is full of malevolent hobgoblins whom it would be better to leave undisturbed. This explains, in large part, why the advisory uses illustrated by cases like *in vino veritas* have gone unnoticed until now. Stalnaker, Heim, and Levinson focused on cases where there's no interesting asymmetry in information regarding the likely consequences of the desire's content.

Nonetheless, in cases where there *is* such an asymmetry, advisory and predictive uses come apart. So we need to tell some story about the also true-sounding but incompatible predictive uses. I offer two possibilities, one more radical than the other. The non-radical proposal posits ambiguity; the more radical proposal attempts to account for predictive uses using the advisory semantic entry, together with general principles concerning assertion. Ultimately, I suggest, the choice between them comes down to empirical questions about the cross-linguistic robustness of advisory uses.

### 9.1 Lexical ambiguity

The ambiguity view is exactly what it sounds like, and doesn't need much explanation. According to it, we simply have two semantic entries for 'wants': one where both the preferences and the

information are hardwired to those of the desirer (Levinson's semantics), and one where the preferences are rigidly indexed to the desirer but the information state is variable. We use one 'wants' to predict what agents will do (in this sense I don't want the Zinfandel) and one to advise them about how to best satisfy their preferences (in this sense I do want the Zinfandel), given the information that's live in the relevant context. This is the view I'd fall back on, if the non-ambiguity view sketched below proves unworkable.

### 9.2 Non-ambiguity

The ambiguity view posits two semantic entries. It seems, at first glance, unavoidable to say something like this. After all, aren't (1) and (4) both true, in different senses, in the *in vino veritas* case, when both uttered and assessed in the same contexts?

Maybe not. It's possible to explain predictive uses, where they differ from advisory ones, with a single advisory entry together with general principles allowing us sometimes to take up the agent's perspective in making assertions. It's not so uncommon an idea that, when we're engaged in the project of explaining and predicting the behavior of agents, we sometimes utter sentences we know to be false, by way of describing the world as it looks from the *agent's* perspective (see Schlenker [2004] for background). Some examples:

- (One police detective to the other, having previously taken the treasure out of the thief's hiding spot): A: "Why is the thief furiously digging there?" B: "He *knows* that the treasure is buried there."
- (In a context where we all know that Achilles hasn't defected to Athens): A: "Why haven't the Trojans invaded Athens yet?" B: "Achilles might have defected to Athens."
- (Said among fellow infidels:) A: "Why is that guy reciting the Athenesian Creed every morning?" B: "If he doesn't, God will smite him."

In none of these cases do we want to use the explanatoriness of the explanans as evidence for fiddling with the semantic entries of their components. In the first case that would give us non-factive knowledge; in the second, a semantics of "might" on which "might $p$" is compatible with "not $p$"; in the last, a theory on which it's fine for atheists to say that God exists and occasionally smites people.

In cases like these, I can successfully explain why someone did something, or predict that they are about to do something, by uttering sentences which I know to be false in my context. I know that it's not true that thief knows that the treasure is buried there. I utter that sentence by way of describing what the thief takes the world to be like, not what the world is actually like. Same with the other two cases: I know that it's not true that Achilles might have defected, because I

know that he didn't defect; but I say that anyway, sketching the world as the Trojans conceive of it, to explain why they're not sending their legions. And describing the world according to the God-fearing man, as if it were actual, can explain why he's muttering the Athenesian Creed each morning.

The non-ambiguity view of the predictive 'wants' holds that the same phenomenon occurs when we use 'wants' to predict and explain agents' actions, in cases where we know that performing that action won't likely satisfy the agent's preferences. In *in vino veritas*, not only *is* there a reading on which (1) is true and (4) is false; that's the only reading that is literally true. There's no sense at all in which I want the Sauvignon Blanc, even if I'm doing everything in my power to buy it, because it's not what will actually satisfy my preferences relative to the information available to those asserting (1). But they can still talk as if I wanted it when they are predicting what I will leave the store with, because I *take* myself to want it. Thus the fine-sounding explanation:

- (Conversation between you and a bystander who also knows that my comrades prefer the Zinfandel): Bystander: "Why is that guy reaching up to that high shelf?" You: "Because that's where the Sauvignon Blanc is, and he wants the Sauvignon Blanc."

On the non-ambiguity view, you just explained my action using a sentence you know to be false in your context. There's not a different entry for 'wants' that tracks what agents *believe* will satisfy their preferences; instead, there's a different kind of speech act, that licenses unembedded quasi-assertions of false sentences, the believing of which makes sense of an agent's behavior.

Is this plausible? To assess this, we'd need a good theory of this general phenomenon, and measure the data we find for 'wants' against it. Here are two relatively flat-footed considerations in its favor. First, it avoids lexical ambiguity, which is always nice when possible. Second, the predictive 'wants' patterns in some key ways like the other dialogues above. One feature paradigmatic of such explanations is that you can coherently continue the dialogue by asserting the negation of the just-seemingly-asserted explanans. In the treasure case, you can coherently continue: "Of course, the thief doesn't *really* know that the treasure is buried there, because it's in our police car." In the Achilles case, you can coherently continue, "Of course, it's not really the case that Achilles might have defected; we all know he didn't." In the God case: "Of course, that's ridiculous; there's no God, and even if there were he wouldn't smite you for forgetting to recite an occasional Athenesian Creed." And—maybe—in the *in vino veritas* case: "Of course, he doesn't *really* want the Sauvignon Blanc, because his comrades prefer the Zinfandel. He really wants the Zinfandel, and someone should go tell him that."

There are, however, some considerations against non-ambiguity. Predictive uses of 'wants' are very common, especially cross-linguistically (see footnote 3). If it turns out that English is unique

in containing advisory uses, that would lend credence to the idea that it has some special word for expressing it. On the other hand, if we find that other languages sometimes contain desire reports whose relevant states of information don't necessarily coincide with the desirer's, that would support a non-ambiguity theory, on which the information state is variable at the level of the semantics. So the choice between these two options may depend on these empirical matters.[17]

## 10 The purpose of desire attributions

It's one thing to give a relativistic semantics for 'wants' that makes some good predictions in cases that make trouble for other semantics. It's another thing to make the case that such a semantics really is plausible. Can it really be that what I want isn't just a function of what the world is like, but also depends on who is attributing the desire to me, and what information *they* have? I want to conclude here with a brief pragmatic sketch of why 'wants' might have evolved in this way (in the spirit of MacFarlane [2014], ch. 12).

What is it to attribute a desire to somebody? One clear answer is the predictive one that I mostly haven't been concerned with here: it's to claim, of that person, that they are psychologically motivated to make the content of that desire come true. If this were the only use we had for attributing desires, we would never utter sentences like (1) in contexts like *in vino veritas*.

But our desires are not all of a piece. We want some things in virtue of wanting other things. I never just want to get on a particular train, end of story; I want to get on that train because it's the train going to Berlin, and I want to go to Berlin. And I don't just want to go to Berlin, either; I want to go to Berlin because that's where my friend is having her birthday party, and I want to be there to help her lament the passing of the years. Plausibly, these chains of explanation eventually bottom out; some things I just *want*, like (maybe) pleasure, or the Good.

The fact that our desires have this kind of structure opens up space for the possibility that you, knowing the general structure of my desires, have access to facts that interfere with these chains of dependence, facts that I myself don't know. Maybe this particular train, which I think I want to get on in order to go to Berlin, *isn't* the train to Berlin, but rather the mislabeled train to Paris. There you are, in the train station bidding me adieu, generally aware of the structure of my desires, just having noticed that I'm about to step on the wrong train.

In a case like this, it makes sense to have a linguistic device to communicate that my preference

[17]Thanks to an anonymous reviewer for pressing this point. Rooryck [2017] points out that, even in English, advisory uses are very hard to hear in the first person. Ambiguity theories can explain this by positing a difference in the two semantic entries for 'wants'. Non-ambiguity theories can account for this too: when I assert, in the present tense, that *I* want the Sauvignon Blanc, the information state parameter is saturated with my information in that context. So the non-ambiguity theory predicts that only a predictive reading is available in such cases.

to go to Berlin stands a much greater chance of being satisfied if I *don't* get on the train I think I want to get on. How are you to get this across? You *could* say, "You ought not get on that train!", but I might misconstrue what you mean. Maybe you've been insisting all along that Berlin is a den of Sin and Debauchery, and have been arguing the whole time that I ought not go to Berlin (even though you are fully aware that, relative to my rather more hedonistic preferences, Sin and Debauchery are things to seek out, not to avoid). What you need is a linguistic device to communicate that you A) are generally aware what my preferences are, and B) have information relative to which they'll actually stand a better chance of being satisfied if I do something other than what I think I want to do. English might have developed any number of such devices, but the one that actually developed is the advisory 'wants'. You yell: "Stop! You don't want to get on that train!" and thereby accomplish exactly your communicative aim.

It's due to this function that 'wants' came to be assessment-sensitive. On the view I've offered, add a claim of the form $\ulcorner x$ wants $\varphi\urcorner$ to the common ground of a conversation is to assert that $\varphi$ outcomes are better than $\neg\varphi$ outcomes, relative to $x$'s base preferences and our best information. Insofar as we care about $x$'s preferences being satisfied, we should strive to make $\varphi$, and not $\neg\varphi$, come true. And if we subsequently acquire more information, information according to which $x$'s preferences now render $\neg\varphi$ better than $\varphi$, we're obliged to take back our prior assertion. To stand by it is to let linger false information about what would be good for $x$ by her lights. That is why it made sense for 'wants' to evolve to be assessment-sensitive; keeping a tally of who wants what is a way of keeping track of what should be done, if we want to help people realize their aims.

## References

Bledin, Justin (2014). "Logic Informed". In: *Mind* 123.490, pp. 277–316.

Bradley, Richard and H. Orri Stefánsson (2016). "Desire, Expectation and Invariance". In: *Mind* 125.499, pp. 691–725.

Buchak, Lara (2013). *Risk and Rationality*. OUP Oxford.

Büring, Daniel (2003). "To want is to want to be there: A note on Levinson 2003". URL: `http://linguistics.ucla.edu/general/Conf/LaBretesche/papers/buring.pdf`.

Callard, Agnes (June 2017). "Everyone desires the good: Socrates' protreptic theory of desire". In: 70, pp. 617–644.

Condoravdi, Cleo and Sven Lauer (Nov. 2016). "Anankastic conditionals are just conditionals". In: *Semantics and Pragmatics* 9.8, pp. 1–69. DOI: `10.3765/sp.9.8`.

Davis, Wayne A. (1984). "The Two Senses of Desire". In: *Philosophical Studies* 45.2, pp. 181–195.

Dowell, Janice J. L. (2011). "A Flexible Contextualist Account of Epistemic Modals". In: *Philosophers' Imprint* 11.14, pp. 1–25.

— (2013). "Flexible Contextualism About Deontic Modals: A Puzzle About Information-Sensitivity". In: *Inquiry* 56.2-3, pp. 149–178.

Egan, Andy (2007). "Epistemic Modals, Relativism and Assertion". In: *Philosophical Studies* 133.1, pp. 1–22.

von Fintel, Kai (1999). "NPI Licensing, Strawson Entailment, and Context Dependency". In: *Journal of Semantics* 16.2, pp. 97–148.

— (2012). "The best we can (expect to) get? Challenges to the classic semantics for deontic modals". In: *Central APA*.

Goble, Lou (1996). "Utilitarian Deontic Logic". In: *Philosophical Studies* 82.3, pp. 317–357.

Heim, Irene (1992). "Presupposition Projection and the Semantics of Attitude Verbs". In: *Journal of Semantics* 9.3, pp. 183–221.

Holliday, Wesley H. and Thomas F. Icard (2013). "Measure Semantics and Qualitative Semantics for Epistemic Modals". In: *Proceedings of SALT 23*, pp. 514–534.

Holliday, Wesley H., Thomas F. Icard, and Matthew Harrison-Trainor (2017). "Preferential Structures for Comparative Probabilistic Reasoning". In: *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*.

Kolodny, Niko and John MacFarlane (2010). "Ifs and Oughts". In: *The Journal of Philosophy* 107.3.

Lassiter, Daniel (2011). "Measurement and Modality: the Scaler Basis of Modal Semantics". PhD thesis. New York University.

Levinson, Dmitry (2003). "Probabilistic model-theoretic semantics for *want*". In: *SALT*. Vol. 13, pp. 222–239.

Lewis, David (1988). "Desire as Belief". In: *Mind* 97.418, pp. 323–32.

MacFarlane, John (2014). *Assessment Sensitivity: Relative Truth and its Applications*. Oxford University Press.

van Rooij, Robert (1999). "Some Analyses of Pro-Attitudes". In: *Logic, Game Theory, and Social Choice*. Ed. by H. de Swart. Tilburg University Press, pp. 263–279.

Rooryck, Johan (2017). "Between desire and necessity: the complementarity of 'want' and 'need'". In: *Crossroads Semantics: Computation, Experiment, and Grammar*. Ed. by Hilke Reckman et al. John Benjamins Publishing Company, pp. 263–279.

Schlenker, Philippe (2004). "Context of Thought and Context of Utterance: A Note on Free Indirect Discourse and the Historical Present". In: *Mind and Language* 19.3, pp. 279–304.

Schroeder, Mark (2011). "Ought, Agents, and Actions". In: *Philosophical Review* 120.1, pp. 1–41.

Stalnaker, Robert (1984). *Inquiry*. Cambridge, MA: MIT Press.

Thomason, Richmond H. (1981). "Deontic Logic as Founded on Tense Logic". English. In: *New Studies in Deontic Logic*. Ed. by Risto Hilpinen. Vol. 152. Synthese Library. Springer Netherlands, pp. 165–176. ISBN: 978-90-277-1346-9. DOI: 10.1007/978-94-009-8484-4_7. URL: http://dx.doi.org/10.1007/978-94-009-8484-4_7.

Villalta, Elisabeth (2000). "Spanish subjunctive clauses require ordered alternatives". In: *SALT*. Vol. 10.

Willer, Malte (2012). "A Remark on Iffy Oughts". In: *Journal of Philosophy* 109.7, pp. 449–461.

Yalcin, Seth (2007). "Epistemic Modals". In: *Mind* 116.464, pp. 983–1026.

— (2012a). "A Counterexample to Modus Tollens". In: *Journal of Philosophical Logic* 41.6, pp. 1001–1024.

— (2012b). "Bayesian Expressivism". In: *Proceedings of the Aristotelian Society* 112.2pt2, pp. 123–160.

— (2012c). "Context Probabilism". In: *Logic, Language and Meaning*. Ed. by M. Aloni. Vol. 7218. Lecture Notes in Computer Science. Springer, pp. 12–21.

## Appendix

Consider the following language:

$$t ::= n_i$$
$$At ::= p_i$$
$$\varphi ::= At \mid \neg\varphi \mid (\varphi \vee \varphi) \mid (\varphi \wedge \varphi) \mid (\varphi \rightarrow \varphi) \mid t \text{ wants } \varphi$$

Let's think of $t$ as a set of names of agents, and $At$ as a set of propositional atoms.

Some abbreviations for readability: let $me = n_1$ with the intended interpretation of me, let $p_z = p_1$ with the intended interpretation of 'my friends prefer the Zinfandel', $p_s = p_2$ for 'my friends prefer the Sauvignon Blanc', $b_z = p_3$ for 'I buy the Zinfandel', and $b_s = p_4$ for 'I buy the Sauvignon Blanc'.

A base model $\mathcal{M}$ for the 'wants'-free fragment of this little language is a pair $\langle W, v\rangle$. $W$ is a finite, non-empty set of worlds, and $v$ is an interpretation function that sends every atom-world pair $\langle w, p\rangle$ to $\{0, 1\}$. Semantic values of formulas are defined relative to models, worlds, and blunt information states. Some definitions:

**Definition.** A **sharp information state** $i$ relative to $\mathcal{M}$ is a pair $\langle S_i, Pr_i\rangle$, where $S_i \subseteq W$ and $Pr_i$ is a function $\mathcal{A} \rightarrow \mathbb{R}_{[0,1]}$, for $\mathcal{A}$ a Boolean algebra of subsets of $W$, such that $Pr_i(S_i) = 1$, and for disjoint $A, B \in \mathcal{A}$, $Pr_i(A \cup B) = Pr_i(A) + Pr_i(B)$.

**Definition.** A **blunt information state** $I$ is a set of sharp information states $i$ such that $\forall i, i' \in I, S_i = S_{i'}$. (Thus we speak without ambiguity of $S_I$.)

**Definition.** $[\varphi]_I = \{w \in S_I : [\![\varphi]\!]^{\mathcal{M},w,I} = 1\}$.

**Definition.** The blunt information state $I$ **updated by** $\varphi$, written $I + \varphi$, is $\{\langle S_i \cap [\varphi]_I, Pr_i^{\varphi}\rangle \mid i \in I\}$, where $Pr_i^{\varphi}(x) = Pr_i(x \mid [\varphi]_I)$.[18]

**Definition.** $[\varphi] = \{w \in W : \forall I,\ [\![\varphi]\!]^{\mathcal{M},w,I} = 1\}$.

**Definition.** A blunt state of information $I$ **accepts** $\varphi$ in $\mathcal{M}$ just in case, for all $w \in S_I$, $[\![\varphi]\!]^{\mathcal{M},w,I} = 1$. In other words, if $[\varphi]_I = I$.

Following Yalcin [2007], Kolodny and MacFarlane [2010], Yalcin [2012a], and Bledin [2014], semantic values of formulas are defined relative to models, worlds, and blunt information states:

$[\![p]\!]^{\mathcal{M},w,I} = 1$ iff $v(w,p) = 1$;
$[\![\varphi \wedge \psi]\!]^{\mathcal{M},w,I} = 1$ iff $[\![\varphi]\!]^{\mathcal{M},w,I} = 1$ and $[\![\psi]\!]^{\mathcal{M},w,I} = 1$;
$[\![\varphi \vee \psi]\!]^{\mathcal{M},w,I} = 1$ iff $[\![\varphi]\!]^{\mathcal{M},w,I} = 1$ or $[\![\psi]\!]^{\mathcal{M},w,I} = 1$;
$[\![\neg\varphi]\!]^{\mathcal{M},w,I} = 1$ iff $[\![\varphi]\!]^{\mathcal{M},w,I} = 0$;
$[\![\varphi \rightarrow \psi]\!]^{\mathcal{M},w,I} = 1$ iff $\forall w' \in S_{I+\varphi},\ [\![\psi]\!]^{\mathcal{M},w',I+\varphi} = 1$.

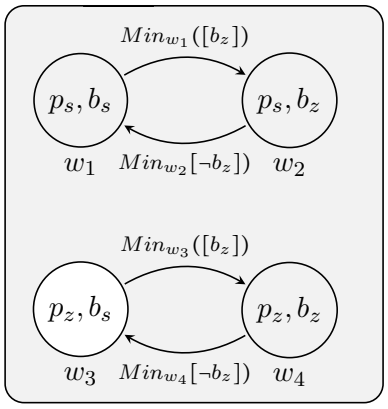
### Heim

A Heim model $\mathcal{M}$ is a tuple $\langle W, Ag, Min, \succeq, B, v\rangle$. $W$ is as before a finite, non-empty set of worlds. $Ag$ is a set of agents. $v$, in addition to assigning semantic values to atoms, assigns members of $Ag$ to agent names $t$. $Min$ assigns to each world $w$ a selection function $Min_w : \mathcal{P}(W) \rightarrow \mathcal{P}(W)$, where $Min_w(A)$ is the subset of $A$ most similar to $w$. $\succeq$ assigns, for each world $w$ and agent $\alpha$, a preorder $\succeq_\alpha^w$ on worlds, representing $\alpha$'s preferences on outcomes $w$. $B$ assigns, for each agent $\alpha$ and world $w$, a set $B_\alpha^w$ of of worlds compatible with the beliefs of $\alpha$ in $w$. As a convention, in models where $\succeq_\alpha^w$ and/or $B_\alpha^w$ do not depend on $w$, we write simply $\succeq_\alpha$ and/or $B_\alpha$, respectively. For sets $W_1$, $W_2$ of worlds, $W_1 \succ_\alpha^w W_2 := \forall w_1 \in W_1, \forall w_2 \in W_2, w_1 \succ_\alpha^w w_2$. With these models, Heim's semantics runs:

$$[\![t \text{ wants } \varphi]\!]^{\mathcal{M},w,I} = 1 \text{ iff } \forall w' \in B_{v(t)}^w, Min_{w'}([\varphi]) \succ_{v(t)}^w Min_{w'}([\neg\varphi])$$

Here's a relatively realistic model of *in vino veritas*. Let's make it a sad model, in which, in the actual world $w_3$, I buy the Sauvignon Blanc, but my friends prefer the Zinfandel. $W = \{w_1, w_2, w_3, w_4\}$, and $v(me) =$ me. Also $Min$ is strongly centered in the model: every $\varphi$ world is its own unique closest $\varphi$ world.

[18] $I + \varphi$ is undefined if $[\varphi]_I$ is empty, which leads to some counterintuitive results and some nice results. One of the counterintuitive ones is that my semantics predicts that you can't want what it's absolutely informationally certain you won't do. But I'm not so concerned with those results here; in paradigmatic instances of the advisory use, namely a context of advice-giving, you think it's not impossible that your advice will be heeded. Causal decision theory, which builds counterfactual notions into the definition of conditional probability, could help here.




$w_1 \approx_{\text{me}} w_4 \succ_{\text{me}} w_2 \approx_{\text{me}} w_3$

The labeled solid lines represent $Min_w$; they point from a world to its closest neighbor(s) in which the label is true. (So the metaphysically most similar world to $w_1$ in which I buy the Zinfandel instead of the Sauvignon Blanc is $w_2$—after all, my decision about which to buy won't affect my comrades' taste.) No basic belief or preference change is included in the model, so we speak of $B_{\text{me}}$ and $\succ_{\text{me}}$. $B_{\text{me}}$ is the entire set—we're thinking of a time before I've made any decisions, so all possibilities are open. Thus at all worlds, all four possibilities are live options in $w$, in the sense important for Heim's semantics: my beliefs don't (yet) rule any of them out.

(1) teaches us that, in a situation like this, ⌜*me* wants $b_z$⌝ should have a true reading. And Heim's semantics does not give us this. This is easy to see:

$$\begin{array}{lll} [\![me \text{ wants } b_z]\!]^{\mathcal{M}, w_3, I} = 1 & \text{iff} & \forall w' \in B_{v(me)}, Min_{w'}([b_z]) \succ_{v(me)} Min_{w'}([\neg b_z]) \\ & \text{only if} & Min_{w_1}([b_z]) \succ_{\text{me}} Min_{w_1}([\neg b_z]) \\ & \text{only if} & \{w_2\} \succ_{\text{me}} \{w_1\} \\ & \text{only if} & w_2 \succ_{\text{me}} w_1 \\ & \text{only if} & \bot \end{array}$$

Heim's semantic entry for 'wants' is not information-sensitive, so it doesn't matter what $I$ is; for any $I$, $w_1$ is a counter-instance to the universal quantifier. Therefore Heim predicts that ⌜*me* wants $b_z$⌝ is false in $w_3$. But it seemed true when you said it in (1)! This is why Heim's semantics fails in predicting the advisory use.

Her semantics also doesn't predict the conditionals, (11) and (12). These conditionals, in our language, are:

(11) $p_z \rightarrow me$ wants $b_z$

(12) $p_s \rightarrow me$ wants $b_s$

The paradigmatic kinds of information states where conditionals like these are asserted are those with open possibilities in which my comrades prefer the Zinfandel, and open possibilities in which

my comrades prefer the Sauvignon Blanc. Thus let $I = \{\langle\{w_1, w_2, w_3, w_4\}, Pr_i\rangle\}$ such that $Pr_i(w) = .25$ for all $w \in S_I$.

$$\begin{array}{rll} [\![p_z \rightarrow me \text{ wants } b_z]\!]^{\mathcal{M}, w_3, I} = 1 & \text{iff} & \forall w' \in S_{I+p_z},\ [\![me \text{ wants } b_z]\!]^{\mathcal{M}, w', I+p_z} = 1 \\ & \text{iff} & [\![me \text{ wants } b_z]\!]^{\mathcal{M}, w_3, I+p_z} = 1 \\ & & \quad \text{and } [\![me \text{ wants } b_z]\!]^{\mathcal{M}, w_4, I+p_z} = 1 \\ & \text{only if} & [\![me \text{ wants } b_z]\!]^{\mathcal{M}, w_3, I+p_z} = 1 \\ & \text{iff} & \bot. \quad \text{(Same calculation as above.)} \end{array}$$

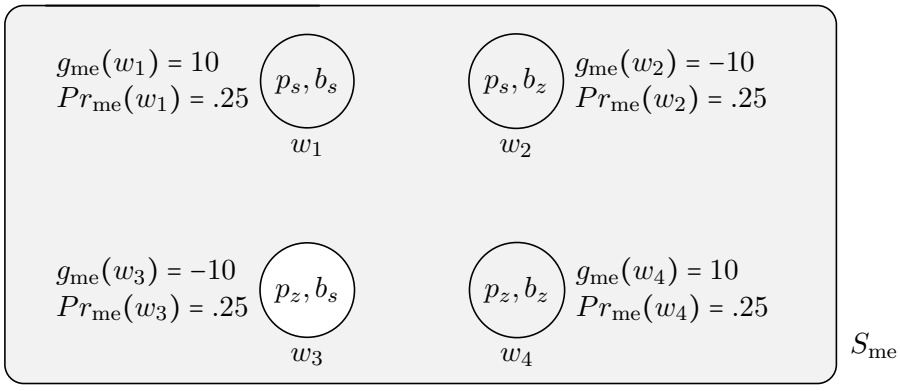
Mutatis mutandis for (12). Thus Heim doesn't predict these conditionals relative to natural models of them, and information states relative to which they are naturally asserted, using a pretty natural semantics for $\rightarrow$.

### Levinson

A Levinson model $\mathcal{M}$ is a tuple $\langle W, Ag, g, Cr, v\rangle$. $W$ is as before a finite, non-empty set of worlds. $Ag$ is a set of agents. $v$, in addition to assigning semantic values to atoms, assigns members of $Ag$ to agent names $t$. $g$ assigns, to each $\alpha \in Ag$ and $w \in W$, a utility function $g^w_\alpha : W \rightarrow \mathbb{R}$. $Cr$ assigns, to each $\alpha \in Ag$ and $w \in W$, a sharp information state $Cr^w_\alpha = \langle S^w_\alpha, Pr^w_\alpha\rangle$, representing that agent's epistemic possibilities and credences. As before, we conventionally drop the world superscripts for $g^w_\alpha$ and $Cr^w_\alpha$, in models where these are stable across worlds. With these models, Levinson's semantics runs:

$$\begin{array}{rll} [\![x \text{ wants } \varphi]\!]^{\mathcal{M}, w, I} = 1 & \text{iff} & EU_{x,w}([\varphi]) > EU_{x,w}([\neg\varphi]) \\ & \text{iff} & \sum\limits_{w' \in S^w_{v(x)}} g^w_{v(x)}(w') Pr^w_{v(x)}(w' \mid [\varphi]) \\ & & \quad > \sum\limits_{w' \in S^w_{v(x)}} g^w_{v(x)}(w') Pr^w_{v(x)}(w' \mid [\neg\varphi]). \end{array}$$

Here is a Levinson model of *in vino veritas*. As before $W = \{w_1, w_2, w_3, w_4\}$, and $v(me) =$ me.




$\ulcorner me \text{ wants } b_z \urcorner$, remember, should have a true reading in a situation like this. What does Levinson's

semantics say about it? Well:

$$\begin{aligned}
[\![me \text{ wants } b_z]\!]^{\mathcal{M},w_3,I} = 1 \quad &\text{iff} \quad EU_{v(me),w_3}([b_z]) > EU_{v(me),w_3}([\neg b_z]) \\
&\text{iff} \quad EU_{\text{me}}([b_z]) > EU_{\text{me}}([\neg b_z]), \quad \text{(Since } B_{\text{me}} \text{ and } g_{me} \text{ are defined in } \mathcal{M}\text{).} \\
&\text{iff} \quad \sum_{w' \in S_{\text{me}}} g_{\text{me}}(w') Pr_{\text{me}}(w' \mid [b_z]) \\
&\qquad > \sum_{w' \in S_{\text{me}}} g_{\text{me}}(w') Pr_{\text{me}}(w' \mid [\neg b_z]) \\
&\text{iff} \quad (10 * 0 + -10 * .5 + -10 * 0 + 10 * .5) \\
&\qquad > (10 * .5 + -10 * 0 + -10 * .5 + 10 * 0) \\
&\text{iff} \quad 0 > 0.
\end{aligned}$$

Since zero is not greater than zero, Levinson's semantics doesn't help.

Same for (11) and (12); relative to the $I$ defined above, the Levinson truth conditions for (12) would run:

$$\begin{aligned}
[\![p_z \rightarrow me \text{ wants } b_z]\!]^{\mathcal{M},w_3,I} = 1 \quad &\text{iff} \quad \forall w' \in S_{I+p_z},\ [\![me \text{ wants } b_z]\!]^{\mathcal{M},w',I+p_z} = 1 \\
&\text{iff} \quad [\![me \text{ wants } b_z]\!]^{\mathcal{M},w_3,I+p_z} = 1 \\
&\qquad \text{and} \quad [\![me \text{ wants } b_z]\!]^{\mathcal{M},w_4,I+p_z} = 1 \\
&\text{only if} \quad [\![me \text{ wants } b_z]\!]^{\mathcal{M},w_3,I+p_z} = 1 \\
&\text{iff} \quad 0 > 0. \quad \text{(Same calculation as above.)}
\end{aligned}$$

Mutatis mutandis for (12). Therefore natural Levinson models of *in vino veritas* do not predict (11) or (12) with respect to natural information states in which they are naturally asserted.

## My proposal

My models are simply Levinson models. The only difference between me and Levinson is the semantic clause for 'wants'.

**Definition.** A mixed expected utility function $Eu^i_g$, relative to a utility function $g$ and sharp information state $i$, is a function: $\mathcal{P}(W) \rightarrow \mathbb{R}$, defined as:

$$EU^i_g(A) \coloneqq \sum_{w' \in S_i} g(w') Pr_i(w' \mid A \cap S_i)$$

My semantic clause for 'wants' is then:

$$\begin{aligned}
[\![t \text{ wants } \varphi]\!]^{\mathcal{M},w,I} = 1 \quad &\text{iff} \quad \forall i \in I, EU^i_{g^w_x}([\varphi]) > EU^i_{g^w_x}([\neg\varphi]) \\
&\text{iff} \quad \forall i \in I, \sum_{w' \in S_i} g^w_{v(t)}(w') Pr_i(w' \mid [\varphi]_I) \\
&\qquad > \sum_{w' \in S_i} g^w_{v(t)}(w') Pr_i(w' \mid [\neg\varphi]_I).
\end{aligned}$$

Relative to the above information state $I$, (11) and (12) both come out true. Here's the derivation of (12). Note that $I + p_s = \{\langle \{w_1, w_2\}, Pr_i^{[p_s]} \rangle\}$, where $Pr_i^{[p_s]}(w_1) = Pr_i^{[p_s]}(w_2) = .5$.

$$\begin{aligned}
[\![p_s \to me \text{ wants } b_s]\!]^{\mathcal{M}, w_3, I} = 1 \quad &\text{iff} \quad \forall w' \in S_{I+p_s},\ [\![me \text{ wants } b_s]\!]^{\mathcal{M}, w', I+p_s} = 1 \\
&\text{iff} \quad [\![me \text{ wants } b_s]\!]^{\mathcal{M}, w_1, I+p_s} = 1 \\
&\qquad \text{and} \quad [\![me \text{ wants } b_s]\!]^{\mathcal{M}, w_2, I+p_s} = 1 \\
&\text{iff} \quad \forall i \in I + p_s, EU^i_{\text{me}}([b_s]) > EU^i_{\text{me}}([\neg b_s]) \\
&\text{iff} \quad 10 * .5 >= -10 * .5 \\
&\text{iff} \quad \top.
\end{aligned}$$

Mutatis mutandis for (11).

Relative to any non-trivial information state $I'$ that accepts $p_z$ (i.e. such that $[p_z]_{I'} = S_{I'}$) and the above Levinson model, my semantics predicts $\ulcorner me \text{ wants } b_z \urcorner$. Any such information state has $S_I = \{w_3, w_4\}$. Thus

$$\begin{aligned}
[\![me \text{ wants } b_z]\!]^{\mathcal{M}, w_3, I'} = 1 \quad &\text{iff} \quad \forall i \in I', EU^i_{\text{me},w_3}([b_z]) > EU^i_{\text{me},w_3}([\neg b_z]) \\
&\text{iff} \quad 5 > -5 \\
&\text{iff} \quad \top.
\end{aligned}$$

The kinds of states of information in which it makes sense to assert the advisory (1)—namely those which accept $p_z$, as in *in vino veritas*—it's true to ascribe to me a corresponding desire to buy the Zinfandel.

Finally, here is an account of consequence modeled after the informational consequence of Yalcin [2012a] and Bledin [2014]:

**Definition.** $\varphi_1 \dots \varphi_n \vDash \psi$ iff, for every $\mathcal{M}$, no information state which accepts $\varphi_1 \dots \varphi_n$ in $\mathcal{M}$ fails to accept $\psi$ in $\mathcal{M}$.

On this definition of consequence, modus ponens comes out valid. Suppose that $I$ accepts $\varphi$ and $\varphi \to \psi$ relative to some arbitrary $\mathcal{M}$. Since $I$ accepts $\varphi$, $I + \varphi = I$. And since $I$ accepts $\varphi \to \psi$, $I + \varphi$ accepts $\psi$. But then $I$ cannot fail to accept $\psi$. Thus modus ponens is valid.

However, modus tollens is not valid. This can be shown using the above model and the information state $I = \langle \{w_1, w_2, w_3, w_4\}, Pr_i \rangle$ where $Pr_i(w_j) = .25$ for all $j$. We've already seen that this information state accepts $p_s \to me$ wants $b_s$. This information state also accepts $\neg(me \text{ wants } b_s)$, for it assigns $b_s$ the same expected utility as $b_z$. However, if the actual world is $w_3$, this model accepts $p_s$. Thus we have a model and and information state relative to which $\varphi \to \psi$ is accepted, $\neg\psi$ is accepted, but $\neg\varphi$ is not accepted.