

# *What does decision theory have to do with wanting?*

Milo Phillips-Brown

Forthcoming in *Mind*

Penultimate draft

*Abstract.* Decision theory and folk psychology both purport to represent the same phenomena: our belief-like and desire- and preference-like states. They also purport to do the same work with these representations: explain and predict our actions. But they do so with different sets of concepts. There's much at stake in whether one of these two sets of concepts can be accounted for with the other. Without such an account, we'd have two competing representations and systems of prediction and explanation, a dubious dualism. Many would be tempted to reject one of the two pictures, yet neither can be let go lightly. Folk psychology structures our daily lives and has proven fruitful in the study of mind and ethics, while decision theory is pervasive in various disciplines, including the quantitative social sciences, neuroscience, and philosophy. My interest is in accounting for folk psychology with decision theory—in particular, for believing and wanting, which decision theory omits. Many have attempted this task for belief. (The *Lockean Thesis* says that there is such an account.) I take up the parallel task for wanting, which has received far less attention. I propose, in the terms of decision theory, necessary and sufficient conditions for when you're truly said to want. I give an analogue of the Lockean Thesis for wanting. My account is an alternative to orthodox accounts that link wanting to preference (e.g. Stalnaker (1984), Lewis (1986)), which I argue are false. I also further that want ascriptions are context-sensitive. My account explains this context-sensitivity, makes sense of conflicting desires, and accommodates phenomena that motivate traditional theses on which 'want' has multiple senses (e.g. all-things-considered vs. pro tanto).

## I Introduction

Folk psychology and descriptive decision theory both purport to represent the same phenomena: our belief-like and desire- and preference-like states. They also purport to do the same work with these representations: explain and predict our actions. That is, both descriptive decision theory and folk psychology aim to show how our belief-like and desire- and preference-like states make sense of how we

---

Thank you for your help: Kyle Blumberg, Marion Boulicault, David Boylan, Alex Byrne, Nilanjan Das, Kevin Dorst, Kai von Fintel, Cosmo Grant, Lyndal Grant, Justin Khoo, Matthew Mandelkern, Robert Pasternak, Ginger Schultheis, Kieran Setiya, Jack Spencer, Mallory Webber, Steve Yablo, two anonymous reviewers, as well as audiences at MIT, the 2019 Pacific APA, Rochester Community and Technical College, the University of Twente, the University of Warsaw, Western Michigan University, and the University of Valencia.

act. You might expect, then, that the concepts of decision theory and those of folk psychology could be accounted for in terms of the other. Can they be?

There is a lot at stake in this question. If its answer were no, then we would be left with a dubious dualism: two competing representations of the same phenomena, two competing systems of prediction and explanation of the same actions. This dualism would tempt many to reject one of the two pictures, yet neither can be let go lightly. Folk psychology structures our daily lives: we understand each other in large part on the basis of notions of folk psychology, such as believing and wanting. These notions have also been fruitful in both the philosophy of mind (e.g. Davidson (1963), Dretske (1988)) and ethics (e.g. Smith (1994)). Decision theory is similarly significant. It holds a pride of place in the quantitative social sciences, especially economics, and is widely used in other disciplines, too, like neuroscience and philosophy.

In this paper, I investigate whether we can account for folk psychology with decision theory. There are two cornerstone notions of folk psychology—believing and wanting—that stand in special need of accounting, for they are omitted by decision theory, which works instead with credence (or degrees of confidence) and preference.

Many have attempted this accounting task for belief, offering necessary and sufficient conditions, stated in terms of credence, for when you're truly said to believe (e.g. Foley (1992), Sturgeon (2008), Easwaran (2016), Leitgeb (2017)).<sup>1</sup> (The *Lockean Thesis* says that there are such conditions.) The same task for wanting has garnered much less attention. It is my task. I give necessary and sufficient conditions, stated in terms of decision theory, for when you're truly said to want;<sup>2</sup> I give an analogue of the Lockean Thesis for wanting.

## 2 The decision-theoretic concept for an account of wanting: expected value

Before considering particular decision-theoretic accounts of wanting, we need to identify the decision-theoretic concepts that we'll use in formulating the accounts.

Decision theory's basic concepts are the credence function, discussed above, and

---

<sup>1</sup>Others argue against these conditions: for example, Ganson (2008), Buchak (2014), Staffel (2016).

<sup>2</sup>There are other things to do, like addressing, as Dreier (2005) has done, Pettit's (1991; 2002) objections that (i) decision theory (putatively) doesn't have the right *desiderative structure* to account for wanting, and (ii) desire (putatively) can take either *prospects* or *properties* as objects (see also (Jackson, 1985)), while decision theory countenances only prospects. Also important is to determine whether desire- and preference-like states are fundamentally one-place (like wanting) or two-place (like preference) (Pollock, 2006).

the value function, which assigns a real number to each outcome.<sup>3</sup> The particular values the value function assigns are in and of themselves immaterial; what matters is the relationship between the values.<sup>4</sup> When you prefer one outcome to another, your value function assigns a higher number to the first. When you prefer one outcome much more than another, or just a little bit more than another, that's represented by the relative distance between the values assigned to those two outcomes. The greater the preference, the bigger the distance.

The credence function alone is of course not what we're looking for—building an account of wanting entirely out of a belief-like notion would be like squeezing blood from a stone. We need a decision-theoretic concept that includes the value function, and decision theory affords us many concepts that do. The one that I'll use is expected value, and I'll use it because of the work that it's done in the theories of others (and the work it does throughout this paper). Expected value, a hybrid of the value function and credence function, is the most commonly used decision-theoretic notion for predicting and explaining action—across a variety of disciplines (Erickson et al., 2013). It's proven fruitful to account for wanting, too. Indeed, expected value has been the concept of choice for all those who've ventured to understand wanting with decision theory.

To get a feel for how wanting and expected value align, consider just one of various things that expected value does: it paints a neat picture of how we want certain things because we believe (or *expect*) they'll lead to other things that we want or value. Two people who at the end of the day value the same things may differ in what they want because they may differ in what they expect to happen. Imagine that O'Neal will give his next paycheck to fight malaria and you and I are debating which organization he should donate to. When all is said and done, you and I value one and the same thing: how many lives are saved. Suppose that O'Neal may donate to the Nets Foundation. I think it's highly likely that the Nets Foundation is effective; you're sure it's a scam. I want O'Neal to donate to the Nets Foundation; you don't.

Because I'm confident that the Nets Foundation is effective, I expect that a donation will likely result in something we both value—many lives being saved—and so I assign a high expected value to the donation. You, though, are sure that the Nets Foundation is a scam, so you expect that saved lives won't result from a do-

---

<sup>3</sup>Just what outcomes are isn't important for my purposes.

<sup>4</sup>Exactly which value function we pick is itself also immaterial. Any function that preserves the relative distances between the values will do.

nation. You assign a low expected value to the donation. I want O’Neal to donate and I assign a high expected value to him donating. You don’t want O’Neal to donate and you don’t assign a high expected value to him donating. This divergence of expected value is no coincidence. You and I diverge in what we want because we diverge in how we expect things to go if O’Neal donates.

Ultimately, expected value may not be quite right for an account of wanting. It has been criticized for its ability to carry out the task of descriptive decision theory: to represent non-ideally rational agents like you and me. Responses have been given, and responses to the responses, and responses to...<sup>5</sup> The decision-theoretic concept for a proper account of wanting will be some entity like expected value—perhaps expected value itself—that emerges from this debate. To keep things simple, I’ll work with expected value.

### 3 Against What’s-best Accounts

We’ve chosen a decision-theoretic concept for our account of wanting—expected value. The rest of this paper is about finding the right expected value-based account.

In this section I criticize what I call *What’s-best Accounts*, accounts that say that you want what’s best in your eyes among a given set of alternatives. The idea that you want what’s best is orthodoxy among those who try to account for wanting in terms of *preference* (e.g. Lewis (1986), Pettit (1991, 2002), Heim (1992), Dreier (2005), Villalta (2008), Condoravdi and Lauer (2016), Rubinstein (2017)). Most What’s-best Accounts on the market aren’t decision-theoretic—few decision-theoretic accounts have been forwarded<sup>6</sup>—but we can translate them into our decision-theoretic framework.

Consider a few What’s-best Accounts.<sup>7</sup>

*S* wants *p* iff *S* prefers *p* to not-*p*. (Davis, 1984)

[W]anting 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, p. 89)

---

<sup>5</sup>Buchak (2013) and Steele and Stefánsson (2016) review the literature.

<sup>6</sup>The decision-theoretic accounts I’m aware of are proposed by van Rooij (1999), Levinson (2003), Wrenn (2010) (whose concern is all-things-considered desire) and Jerzak (2019).

<sup>7</sup>‘*S*’ ranges over the names of agents; ‘*S*’ ranges over the corresponding agents; ‘*p*’ ranges over proposition-denoting strings; and (ignoring any context-dependence in *p*) ‘*p*’ ranges over the corresponding propositions.

⌈S wants  $p$ ⌋ is true at  $w$  iff  $p$  is true in all of the *best* worlds compatible with  $S$ 's beliefs at  $w$ , as ranked by  $S$ 's preferences at  $w$ . (paraphrasing von Fintel (1999))

⌈S wants  $p$ ⌋ is true iff the expected value  $S$  at  $w$  assigns to  $p$  exceeds the expected value  $S$  assigns to not- $p$ . (van Rooij, 1999; Levinson, 2003)

(Most of these accounts presuppose that wanting is a propositional attitude, a pedigree, if contested (e.g. Montague (2007)), presupposition. To keep things simple, I'll take on this presupposition too.)

Here's a view, within the expected value framework, that captures the basic idea of the accounts just above.

*Simple What's-best Account*

⌈S wants  $p$ ⌋ is true at  $w$  iff  $S$  at  $w$  assigns a higher expected value to  $p$  than to any of certain alternatives.

(Levinson, for example, says that there's just one certain alternative: not- $p$ .<sup>8</sup>)

The Simple What's-best Account is false. Its problem is shared by its more sophisticated What's-best Account brethren, decision-theoretic and otherwise: being best is neither necessary nor sufficient for being wanted. Certain mundane facts about wanting bring the problem into view.

Being best is not sufficient for being wanted because sometimes we want none of the options we're faced with, even the best one—a fact that has gone unacknowledged by advocates of What's-best Accounts. Imagine that you have been kidnapped and must make an awful choice: either shoot one of the two people in front of you, or do nothing and both will be shot. Not being a sociopath, you neither want to shoot, nor do you want to refrain from shooting and have the two be shot! Although it's not true that you want to shoot, shooting is nonetheless best: you prefer shooting the one to not shooting the one and having both be shot. Shooting is best, but you don't want it.<sup>9</sup> (Maybe you can be truly said to want to

---

<sup>8</sup>For Crnić (2011) and Phillips-Brown (2018), the certain alternatives are indexed to a partition on the set of possible worlds, paralleling Lockean Theses—e.g. Locke's (2014) and Leitgeb's (2017)—on which belief is partition-relative. (For a partition-relative view of belief outside of the Lockean literature, see e.g. (Yalcin, 2016).)

<sup>9</sup>Advocates of What's-best Accounts do have a possible reply. There are two parts to my claim that being best is not sufficient for being wanted: first, that you do not want to shoot the one, and second, that the alternative to shooting the one is not shooting and letting the two be killed, which makes shooting the one best. The reply would be to accept the first part but deny the second. In particular, to say that not shooting and letting the two be killed is not the only alternative; rather, there's some additional alternative that you prefer to shooting the one, making shooting the one *not* best. If you like this reply,

shoot, on which more in section 8. But you can also clearly be truly said not to: that's the problem for What's-best Accounts.)

Or suppose that you are deeply, deeply depressed. There is *nothing at all* in the whole world that you want. Life is misery. Even so, you do prefer some things to others. Something is best, but nothing is wanted.

Being best is not necessary for being wanted because sometimes we want many things, even if one of them isn't best. Suppose that you're going out to dinner. Your options are the pizzeria, the ramen shop, and the hot dog stand, and while hot dogs sound bad tonight, the other two options sound good. The pizzeria would be best. You want to go the ramen shop, and you want to go to the pizzeria even more. You want to go to the ramen shop, but you *disprefer* it to one of the other alternatives. You want it, but it's not best.

Proponents of What's-best Accounts—in particular, Levinson (2003) and Crnić (2011)—have claimed that cases like our pizza–ramen one are not in fact counterexamples to the thesis that being best is necessary for being wanted. They hypothesize that 'want' is *context-sensitive*. We should represent agents with many value functions, not just one, and which value function we use to evaluate a want ascription differs by context. In each context, what's wanted is what's best according to the value function in that context.

The different value functions are supposed to represent different dimensions of value that matter to the agent. For example, one function will represent your value of eating things that they are bready and cheesy, assigning higher values to outcomes to the extent that she eats things that are bready and cheesy in them. Relative to this function, pizza is best, while ramen is best according to a function that favors eating brothy soups with noodles.

More formally, the idea is as follows. Where  $c$  is a context:

*What's-best Account with Varying Value Functions*

$\lceil S$  wants  $p \rceil$  is true in  $c$  at  $w$  iff  $S$  at  $w$  assigns a higher expected value to  $p$  than to any of certain alternatives, *relative to the value function in  $c$ .*

*Formally*, this does save the thesis that being best is necessary for being wanted. For example, we need both of the sentences on the following page to be true:

---

then the shooting case is not a counterexample. Nonetheless, What's-good-enough Accounts are still committed to *something* being wanted in all cases, since in all cases, something is best. We still have counterexamples, then, in cases where nothing is wanted, like the depressive case just below.

*Pizza.* You want to go to the pizzeria.

*Ramen.* You want to go to the ramen shop.

Both can be true if *Pizza* is evaluated in a context with a value function that rates the pizzeria as best, and if *Ramen* is evaluated in a different context, one whose value function ranks the ramen shop best.

The varying value functions don't merely help formally in this case. What they're supposed to represent looks right too. The value function against which *Pizza* comes out true represents your interest in things bready and cheesy, while *Ramen* comes out true against a value function that represents your interest in things noodley and brothy. That feels right.

There are cases, though, where the account would have us posit multiple dimensions of value where there is only one. Imagine that you will be given a single ticket from a hat. Most of the tickets are worthless. Two tickets, though, have cash value, the red ticket (worth 50 dollars) and the blue ticket (worth 100 dollars). You want to get the red ticket, and of course you also want to get the blue ticket. Formally, the two sentences would be true only if there are two contexts, each with a different value function. Those value functions are supposed to represent different dimensions of value. Along one dimension, the red ticket is better than the blue ticket; along the other it's flipped. But what would these different dimensions of value be? The tickets are both good for the same thing: money. And as far as money goes, 100 dollars is unambiguously better than 50; the blue ticket is unambiguously better than the red. The dimension on which the red ticket is valued (money) is not one on which it's best. The fact that you both want the red ticket and want the blue ticket can't be traced to multiple dimensions of value.

Let's put What's-best Accounts to rest.

#### 4 A Simple What's-good-enough Account

What's-best Accounts are misguided. An alternative approach comes from what I call *What's-good-enough Accounts*, of which my own account is one. (Only three other accounts that fall under the 'What's-good-enough Accounts' umbrella have been proposed—none are motivated in quite the way that I motivate mine. van Rooij's (1999) account I note just below; the others I discuss in footnote 15.)

What's-good-enough Accounts claim that you want, well, what's good enough (in your eyes). This rights the wrongs of What's-best Accounts. Recall that you

want to go to the ramen shop and want to go to the pizzeria more. Why is it that you can want two things (even though one of them isn't best)? Intuitively, because both are good enough. You'd be happy going to the pizzeria and happy going to the ramen shop; either will do. Recall also that you want neither to shoot the one nor refrain from shooting with the other two ending up shot. Why is it that you can want neither of the two things (even though one of them is best)? Intuitively, because neither is good enough. You don't want to shoot the one or refrain from shooting, because doing either would result in something truly awful.

A Simple What's-good-enough Account is below. Something is good enough for an agent when she assigns it an expected value that meets a certain threshold, a real number.

*Simple What's-good-enough Account*

⌈S wants  $p$ ⌋ is true at  $w$  iff the expected value  $S$  at  $w$  assigns to  $p$  meets a certain threshold.

(van Rooij (1999) suggests that the threshold is the expected value that  $S$  at  $w$  assigns to a tautology.<sup>10</sup>) Compare the Simple What's-good-enough Account to a simple version of the Lockean Thesis: ⌈S believes  $p$ ⌋ is true at  $w$  iff the credence  $S$  at  $w$  assigns to  $p$  meets a certain threshold.

You want to go to the ramen shop and you want to go to the pizzeria. The account can make sense of both facts by saying that the expected value you assign to going to each place meets the threshold. You neither want to shoot the one nor refrain from shooting, and the account can accommodate both of these facts, too—this time by saying that neither option meets the threshold.

What's-good-enough Accounts have another thing going for them: they neatly explain the pervasive phenomenon of wanting  $p$  while simultaneously wanting not- $p$  (Baker (2010) calls these *directly conflicting desires*). Imagine that your daughter is deciding whether to take over the family business. You both want her to take over (it's a generations-old tradition) and want her not to (it would be good for her to find her own way in life). This has proved puzzling from a theoretical perspective (e.g. Jackson (1985); Ashwell (2017)). *Believing* both that your daughter will take over the business and that she won't is paradigmatically irrational; more generally it's irrational to both believe  $p$  and believe not- $p$ . But, intuitively, the same doesn't hold for wanting. Why should that be?

---

<sup>10</sup>For Pollock (2006), you want something iff it's both best *and* good enough.



On What's-good-enough Accounts, cases of directly conflicting desires couldn't be simpler: both  $p$  and not- $p$  can be good enough, and so both can be wanted. Both your daughter taking over the business and her not doing so can be good enough; both can be wanted.

The fact that What's-good-enough Accounts neatly explain directly conflicting desires undermines an objection that's been made against the broader project of accounting for wanting with preference. (Baker, 2010) is dedicated to establishing that the project is doomed by directly conflicting desires. But his case rests on a false presupposition: that preference-based accounts of wanting are What's-best Accounts. (His statement of preference-based accounts is: 'that an agent wants  $p$  reduces to preferring  $p$  to certain potential alternatives' (p. 42).) He rightly argues that What's-best Accounts cannot account for directly conflicting desires. (A What's-best Account would say that if you want  $p$  and simultaneously want not- $p$ , then you prefer  $p$  to not- $p$  and simultaneously prefer not- $p$  to  $p$ , which *is* irrational.) But of course we can adopt a What's-good-enough Account instead, and if we do, directly conflicting desires needn't be problem.

What's-good-enough Accounts are, I believe, the right kind of account. But the Simple What's-good-enough Account itself is wrong, as we'll see in the next section.

## 5 Interlude: the context-sensitivity of what you're truly said to want

We have to pause our discussion of just how wanting relates to decision theory. To give any such account of this relation, *or any account of wanting at all*, we must appreciate a certain fact: what you're truly said to want isn't intrinsic to you; it's partly determined by ascribers, and varies by context.

Add a further detail to our dinner case: I will be driving you to your dinner destination. I'll turn right for the pizzeria, left for the ramen shop. I ask where you want to go. I can't both turn left and right; we can't go to both the pizzeria and the ramen shop. Recall that you prefer the pizzeria to the ramen shop. Before, you truly uttered *Pizza* and *Ramen*. But you must now commit. Make up your mind, I'd say, which do you want? You must reply with what you prefer most, given the choice between the three dinner options. What you prefer most, remember, is the pizzeria, so *Pizza* is true. Because you disprefer the ramen shop, *Ramen* is false in this new context. (Or, to bring out the point another way, imagine an on-looker hearing you ask me to drive to the pizzeria. He could say, 'I guess she [you] ultimately does not want to go to the ramen shop'. In this context, *Ramen* is false.<sup>11</sup>)

In the last section, there was a context where *Ramen* is *true*. But now we've just seen a new context, a context with no change in your psychological state—no change intrinsic to you—where *Ramen* is *false*. What you're truly said to want is not intrinsic to you; it's context-sensitive.<sup>12</sup>

The context-sensitivity of what you're truly said to want falsifies any account of wanting, and I mean any account, *decision-theoretic or otherwise*, that ignores context when giving necessary and sufficient conditions for wanting. Such accounts, the Simple What's-good-enough Account among them, simply say that you're truly said to want *p* just if some condition *C* obtains, where *C* is insensitive to context. To take just some examples:

*Simple What's-best Account* (repeated from above)

⌈*S* wants *p*⌋ is true at *w* iff *S* at *w* assigns a higher expected value to *p* than to any of certain alternatives.

*Simple What's-good-enough Account* (repeated from above)

⌈*S* wants *p*⌋ is true at *w* iff the expected value *S* at *w* assigns to *p* meets a certain threshold.

*S* wants *p* iff *S* is disposed to take pleasure when it seems to her that *p* obtains.<sup>13</sup>

*S* wants *p* iff *S* is disposed, other things equal, to do what she believes will bring it about that *p* obtains.<sup>14</sup>

*S* has an 'intrinsic appetitive desire that *P*' iff *S* 'constitute[s] *P* as a reward'. (Arpaly and Schroeder, 2013, p. 128)

Read flatfootedly, the conditions *C* that these accounts place on wanting—the right hand side of the biconditionals—don't reference context. For example, the account that concerns pleasure does *not* say ⌈*S* wants *p*⌋ is true in *c* iff *S* is disposed to take pleasure when it seems to her that *p* obtains, given some constraint imposed by *c*. Without mention of context, a contradiction follows. Given that *Ramen* is true (in one context), the accounts say that *C* (taking pleasure in a certain thing, being disposed to act in a certain way, etc.) obtains. But they also say that *C* does *not* obtain, since *Ramen* is false (in another context).

---

<sup>12</sup>Others offer other data that suggest context-sensitivity: e.g. Villalta (2008), Lassiter (2011), and Condoravdi and Lauer (2016).

<sup>13</sup>Morillo (1990) advances such a view.

<sup>14</sup>This is orthodoxy (Ashwell, 2017).

I'm not claiming that there's something faulty about these basic *approaches* to wanting—in terms of pleasure, dispositions to act, etc.—but rather that their instances here aren't quite complete. It's no great mystery how to fix them: the condition C needs to be constrained by context, so that the same want ascription can be true in one but false in another. Note, though, that this will change the character of the accounts, each of which, as written, makes what you're truly said to want *intrinsic* to you: that you're disposed to take pleasure in a certain thing, that you're disposed to act in a certain way, etc. Yet the constraints that context places on the condition C are partly under the control of ascribers; they don't depend entirely on what's going on inside of you. Adding these constraints makes what you're truly said to want not intrinsic to you.

It might strike you that there's a simple fix to the problem of context-sensitivity for want ascriptions. Many have hypothesized that 'want' has exactly two *senses*—in other words, that 'want' is (two-way) ambiguous. For example, 'want' has been thought to express either pro tanto or all-things-considered desire; a pro attitude or a volitive attitude (Daveney, 1961); an appetitive attitude or a volitive attitude (Davis, 1984). On such hypotheses, you'd expect that *Ramen's* truth value would be slippery. *Ramen* would be true with one sense and false with another.

Even if there are multiple senses of 'want', and I'll suggest that there aren't in section 8, there can't be exactly two. Add yet another detail to the case. You are a gourmet, and the food options where you live aren't up to your ideal. You say, 'of course neither of these places do I *want* to go to. (They don't hold a candle to what's in New York, for example. It's just that in this culinary wasteland, they're the only places that pass for decent.)' In this new context, both *Pizza* and *Ramen* are false. We have three contexts on our hands: first the one where both *Pizza* and *Ramen* are true; second the one where *Pizza* is true but *Ramen* is not; and now, third, the one where both are false. Three contexts with shifting truth values are one more than a hypothesis of exactly two senses can handle.

## 6 A context-sensitive What's-good-enough Account

The goal is to give a decision-theoretic account of when you're truly said to want, and that, we now know, shifts by context.

I believe that what's shifting by context is *what counts as good enough*. Or, formally, what shifts is the threshold. A first pass implementation of this idea, to be refined in the next section, is this:

*What's-good-enough Account with Context-sensitive Threshold*

$\lceil S \text{ wants } p \rceil$  is true in  $c$  at  $w$  iff the expected value  $S$  at  $w$  assigns to  $p$  meets the threshold in  $c$ .<sup>15</sup>

Compare to a context-sensitive version of the Lockean Thesis:  $\lceil S \text{ believes } p \rceil$  is true in  $c$  at  $w$  iff the credence  $S$  at  $w$  assigns to  $p$  meets the threshold in  $c$ . (This sort of context-sensitivity is more or less standard now among Lockeans about belief; see e.g. Sturgeon (2008), Leitgeb (2017).) Further, note that on the What's-good-enough Account with Context-sensitive Threshold, 'want' is context-sensitive. This aligns with a more general project according to which attitude verbs—among them 'believe', 'know', 'surprise', and 'suspect'—are context-sensitive.<sup>16</sup>

The What's-good-enough Account with Context-sensitive Threshold inherits the Simple What's-good-enough Account's merits, and as we'll now see, corrects its failure in cases of context-sensitivity.

Call to mind our three contexts: the one where *Pizza* and *Ramen* are both true, the one where just *Pizza* is true, and the one where neither are. Formally, the What's-good-enough Account with Context-sensitive Threshold says that in the context where both are true, the threshold is met both by the expected value that you assign to the pizzeria and by the expected value that you assign to the ramen shop. Where *Pizza* is true but *Ramen* is false, the threshold is higher, met by the pizzeria but not the ramen shop. The threshold reaches even greater heights in the context where both sentences are false. Neither the ramen shop nor the pizzeria meet it. All of our cases are covered.<sup>17</sup>

<sup>15</sup>Bradley's 1999 and Lassiter's 2011 fall into this bucket. Bradley, who's concerned with conditional desire, says in passing, and without elaborating, 'to desire that  $X$  is simply to desire  $X$  more than the status quo or whatever other threshold is assumed in a particular context' (p. 26).

For Lassiter, what's good enough is *what's significantly greater than average*. Average is represented by  $Alt_c(p)$ , a contextually determined set of propositions that includes  $p$ . (Lassiter's view thus bears some resemblance to the partition-relative views from footnote 8.)  $\lceil S \text{ wants } p \rceil$  is true in  $c$  iff the expected value  $S$  assigns to  $p$  is 'significantly greater' (p. 182) than the expected value  $S$  assigns to  $\bigcup Alt_c(p)$ . Being significantly greater than the average, though, is neither necessary nor sufficient for being wanted.

*Not necessary.* Assume that to the pizzeria you assign value 6; ramen shop, 2; hot dog stand, -2. There's an equal chance, suppose, that you'll be taken to each:  $\bigcup Alt_c(\text{ramen shop})$ 's expected value is 2. You want to go to the ramen shop, but the expected value you assign it (2) isn't significantly greater than average (also 2).

*Not sufficient.* Ani is vegetarian and dangerously allergic to shellfish. She'll be served, with equal likelihood, ratatouille (value 6), chicken (-30), or lobster (-900). The expected value she assigns to chicken is thus -30. That's significantly greater than average, -308, yet it's false that Ani wants chicken.

<sup>16</sup> On 'believe' outside of the Lockean literature see e.g. Stalnaker (2008). See e.g. Stine (1976) on 'know'; Villalta (2008) on the analogues of 'fear', 'hope', and 'glad' in Spanish; Blumberg and Holguín (2019) on 'surprise' and 'suspect'.

<sup>17</sup>There's a complication here, as an anonymous reviewer points out. Recall the case where I as your driver ask where you want to go. Right for the pizzeria, left for the ramen shop. You say:

Let's explore more what it means for something to count as good enough, a notion that we've so far been understanding on an intuitive level. What counts as good enough, I believe, is determined by the communicative purposes of ascribers, and what counts as good enough shifts by context because communicative purposes shift by context. I will venture hypotheses about where the threshold is set in particular cases. These hypotheses are not supposed to be definitive, but rather make up a sort of proof of concept of the idea that the threshold is fixed by communicative purposes.

Start with the context where you say (repeating now from above), 'Of course neither of these places do I *want* to go to. (They don't hold a candle to what's in New York, for example. It's just that in this culinary wasteland, they're the only places that pass for decent.)' Here, what counts as good enough is going to a

- 
- (i) a. I don't really want [i.e. it's not the case that I want] to go to the pizzeria,
  - b. but yeah, I want you to turn right here.

*If* (i) is evaluated in a single context, there's potential trouble for the What's-good-enough Account with Varying Threshold—on which the conjuncts of (i) contradict each other in a single context. Yet some hear (i) as true! (Why the contradiction? Imagine you assign the same expected value,  $V$ , to turning right and to going to the pizzeria: the threshold in a single context is then simultaneously met by  $V$  (by (i-b)) and not (by (i-a)).) In fact, everything's copacetic; (i-a) and (i-b) are sometimes evaluated in *different* contexts, or so I claim. While some hear (i) as true, *others bear it as false*, and when a sentence is heard as true by some and false by others, a difference in context is often why. That's *prima facie* reason to accept my claim; I'll now argue for it in detail.

Consider two sentences that share truth conditions with (i):

- (ii) a. I don't want to go to the pizzeria,
- b. and I want you to turn right here.
- (iii) a. I don't really want to go to the pizzeria (it's just not that good),
- b. but the options in town being what they are, yeah, I want you to turn right here.

(ii) makes no sense, while (iii) sounds fine. Why the difference? In any such situation—two sentences with the same truth conditions with different judgments—the usual suspect is context change, and it's indeed the culprit here. There's context change within (iii) but not within (ii). The difference in context change traces to a difference in how much material the sentences contain. (iii) is just (ii) with extra material (and 'but' instead of 'and'), which suggests a change in communicative purposes between its conjuncts, and so a change in context. In (iii-a), 'it's just not that good' suggests that the communicative purpose of 'I don't really want to go to the pizzeria' is to indicate displeasure at the quality of the pizzeria. In (iii-b), 'but the options in town being what they are' suggests that the communicative purpose of 'I want you to turn right here' is to indicate to your driver that you most prefer turning right. (ii) has no extra material; thus, no change between conjuncts in communicative purpose or context.

There's a spectrum of judgments: (ii) makes sense to none, (i) to some but not all, (iii) to all. Correspondingly, there's a spectrum of material: (ii) the least, (i) is in the middle; (iii) the most. This match of spectra is no accident. The extent that each (i)–(iii) make sense matches the extent that we perceive communicative purposes—and so contexts—as changing, and how we perceive communicative purposes is driven by the extent of extra material. We've already seen how this works with (ii) and (iii). I hypothesize that the reason that (i) makes sense to some, but not others, is that some, but not others, glean enough from the extra material in (i) to tell a story about why communicative purposes would change between the conjuncts of (i). (Certainly, this is true of me when I hear (i) as true.) Those who don't hear (i) as true don't glean enough.

restaurant that, to your mind, is of a certain quality—that of an average restaurant you like to go to in New York, say. Neither the pizzeria nor the ramen shop compare favorably to that kind of restaurant, which is to say they don't meet the threshold. That's why *Pizza* and *Ramen* are both false in this context.

A similar thing happens in the context where you can either shoot the one or two will be shot. Plausibly, the reason that you speak truly in saying that you neither want to shoot nor refrain from shooting is that you're implicitly comparing the choices available to you in this situation you've been forced into to the choices that would be available if you weren't under duress. (Imagine saying, 'I don't want to shoot and I don't want to not shoot and let the two be shot! But I'm being *forced* to do one!') When what counts as good enough are actions you'd have available in a situation you would choose to be in, neither shooting nor not shooting and letting the two be shot is good enough. Put differently, both shooting and not look horrible in comparison to the kinds of actions you'd choose to have available.<sup>18</sup>

Recall the thought that folk psychology serves to predict and explain our actions. In communicating to your interlocutor that you prefer big city restaurants to the ramen shop and the pizzeria, you give her information on which to predict your future behavior: when given greater choices, you wouldn't choose the ramen shop or the pizzeria, even though in your current circumstance, you would. In communicating that you disprefer shooting to what you would do if you weren't under duress, you position your interlocutor to predict how you'd act if you weren't under duress.

Think now about the case where I am your driver and, asking you where you want to go, will turn the car based on your answer. As we've said, what you're truly said to want here, among the dinner options, is only the one that you most prefer. Part of our communicative purpose in this case is for you to instruct me where to go; it would stand to reason that you should be instructing me to go only where you most prefer to go, which is why what counts as good enough is only what you most prefer.<sup>19</sup>

More generally, we have an insight into cases where what counts as good enough is only what you most prefer, given certain options. Ascribing a desire to someone

---

<sup>18</sup>Daveney (1961) discusses how 'want' can communicate how one would choose to act when under duress versus not.

<sup>19</sup>Things are more complicated when there's a tie in what you most prefer. I'll leave this issue for another day.

in such a context again gives the audience valuable information. Given that agents tend to do what they most prefer to do—excepting cases of weakness of will or irrationality—the audience can predict that the agent will try to do what’s she been said to want to do in these contexts. And if the agent ends up acting in that way, the audience can explain her action by pointing to the fact that what she did was what she most preferred to do.

Finally, take the case where you truly say that you want to go to the pizzeria, that you want to go to the ramen shop, and that you don’t want to go to the hot dog stand. Suppose that when you assert all this, you and your conversational partner are negotiating on where to go to dinner. I hypothesize that what counts as good enough is anything Orange that’s preferable to cooking at home; or, perhaps, what counts as good as enough is anything significantly preferable to cooking at home—i.e. the threshold is set significantly higher than the expected value that you assign to cooking at home. It’s helpful for the communicative purpose at hand for the context to be set in this way. You are saying, more or less, that you’re willing to go to dinner at either the ramen shop or the pizzeria, but unwilling to go to the hot dog stand.

## 7 An upgraded context-sensitive What’s-good-enough Account

According to What’s-good-enough Accounts, you’re truly said to want what counts as good enough, where what’s good enough is represented by a threshold. The innovation from the last section is that this threshold varies by context. It turns out—as we’ll see in this section—that the threshold varies not just by context, but also by the world at which we evaluate a want ascription, the world of evaluation. Put another way, the threshold is determined partly by the world of evaluation. This fact falsifies the What’s-good-enough Account with Context-sensitive Threshold (from the last section), which says that the threshold is determined only by the context. We need to upgrade this account to one where the threshold is determined by both context and the world of evaluation—the upgraded account is my view.<sup>20</sup>

To see that the threshold can vary with the world of evaluation, consider the following case. Amelia is on a plane and wants to watch a movie, despite the fact that the movie selection is what you’d usually find on a plane: not great. The best movie on offer is *Hotel Transylvania*. Fix a context: one where you’re truly said to

---

<sup>20</sup>Thank you to one anonymous reviewer for raising the issue outlined in this paragraph (it prompted me to write this section) and to another for helping me make my presentation clearer.

want what's best among your available options. Amelia's traveling companion says:

- (1) a. Amelia wants to watch *Hotel Transylvania*,
- b. but of course she wouldn't want to if *Jaws* were available.

To know how to evaluate (1-b), we need a semantics for counterfactual conditionals on the table. For concreteness, take Stalnaker's (1968): a counterfactual  $\lceil \text{if } r, q \rceil$  is true in  $c$  at  $w$  iff  $q$  is true in the closest  $r$ -world to  $w$ . In other words, given a context  $c$ ,  $\lceil \text{if } r, q \rceil$  is true at a world of evaluation  $w$  iff its consequent,  $q$ , is true at a different world of evaluation—namely, the closest  $r$ -world to  $w$

(1) is true in the actual world, suppose. And so because (1-b) is actually true, it's not the case that Amelia wants to watch *Hotel Transylvania* at the closest world to actuality where *Jaws* is available. But of course in the actual world, Amelia *does* want to watch *Hotel Transylvania*, since (1-a) is actually true.

The situation, then, is this. Where  $w_{@}$  is the actual world and  $w_{Jaws}$  is the closest world where *Jaws* is available: in a single context, it's true in one world of evaluation,  $w_{@}$ , that Amelia wants to watch *Hotel Transylvania*, but not true at another world of evaluation,  $w_{Jaws}$ . (And it's at  $w_{Jaws}$  that we must evaluate the consequent of (1-b).) Watching *Hotel Transylvania* is good enough at  $w_{@}$ , but not at  $w_{Jaws}$ .

Why this difference between worlds? I hypothesize that it's because there's a difference between worlds in what counts as good enough.<sup>21</sup> We're in a context where what counts as good enough is the best of the available options, and the best available option at  $w_{@}$  (*Hotel Transylvania*) differs from the best available option at  $w_{Jaws}$  (*Jaws*). What counts as good enough is, of course, represented by the threshold. There's a difference in threshold between worlds without any difference in context: the threshold depends partly on the world of evaluation.

The world of evaluation plays a similar role in other cases, too. In the last section, for instance, we considered a context where what counts as good enough among the dinner options is anything that's preferable to cooking at home or (perhaps) anything significantly preferable to cooking at home. How good it is to have dinner at home differs by world (in some worlds, you have filet mignon in the fridge, in others, week-old leftovers), and in turn, what counts as good enough differs by world—even though the context stays the same.

---

<sup>21</sup>Might the difference instead trace to a difference in the expected value Amelia assigns to seeing *Hotel Transylvania* between worlds? It doesn't; or anyway it needn't. Imagine that in these two worlds, *Hotel Transylvania* has the same cast, the same score, the same plot, the same everything. In both worlds, Amelia assigns the same expected value to seeing it.



A difference in threshold absent a difference in context falsifies the What’s-good-enough Account with Context-sensitive Threshold, on which  $\lceil S \text{ wants } p \rceil$  is true in  $c$  at  $w$  iff the expected value  $S$  at  $w$  assigns to  $p$  meets the threshold in  $c$ . On this account, the threshold is determined by context alone.

We need an upgraded account, one on which the context and the world of evaluation together determine the threshold. How *do* they determine the threshold? I believe that context plays a leading role. It tells us how to get from the world of evaluation to the threshold. Take the context where what counts as good enough is the best of the available options. Context says, give me a world  $w$  and I’ll tell you the threshold to use when evaluating a want ascription at  $w$ : the threshold is the expected value that you at  $w$  assign to the best of the available options at  $w$ .

Put formally, what I’m proposing is that context determines a function—a *threshold profile*—from worlds to thresholds. Where  $t_c$  is the threshold profile of  $c$ , I submit:

*What’s-good-enough Account with Context-sensitive Threshold Profile*<sup>22</sup>

$\lceil S \text{ wants } p \rceil$  is true in  $c$  at  $w$  iff the expected value  $S$  at  $w$  assigns to  $p$  meets  $t_c(w)$

We have the familiar idea of the What’s-good-enough Account: you’re truly said to want  $p$  just if the expected value you assign to  $p$  meets the threshold. What’s new is that we’ve upgraded our theory on how the threshold is determined—it’s determined by both the context and the world of evaluation.

The upgrade enables us to neatly get the right predictions in the movie case. In that case, the context  $c$  is one where what counts as good enough (in a world) is the best of the available options (in that world). The What’s-good-enough Account

---

<sup>22</sup>We actually need the threshold to also be (partly) determined by the *agent*. Consider why an agent-independent threshold, like those in the main text, won’t do. It’s contested whether interpersonal comparisons of utility are possible—i.e. whether we can sensibly compare the value functions of two agents (Binmore (2009) reviews the literature). If we *can’t*, then we can’t sensibly compare the two expected values that two agents assign to some proposition to a single number, or threshold: we need one threshold for each agent. And even if we *can*, there’s still a problem. Take the context where one is truly said to want the best of the available options: the What’s-good-enough Account with Context-sensitive Threshold Profile says  $\lceil S \text{ wants } p \rceil$  is true in  $c$  at  $w$  iff the expected value  $S$  at  $w$  assigns to  $p$  meets  $t_c(w)$ —the expected value  $S$  at  $w$  assigns to the best of the available options. For another agent,  $S'$ ,  $\lceil S' \text{ wants } p \rceil$  is true in  $c$  at  $w$  iff the expected value  $S'$  at  $w$  assigns to  $p$  meets  $t_c(w)$ —the expected value  $S'$  at  $w$  assigns to the best of the available options. But the two agents needn’t assign the same expected value to this option! The threshold can’t be given by  $t_c(w)$ , but instead must be made relative to the agent.

I propose a function,  $t_c^+$ , from worlds *and agents* to thresholds.  $\lceil S \text{ wants } p \rceil$  is true in  $c$  at  $w$  iff the expected value  $S$  at  $w$  assigns to  $p$  meets  $t_c^+(w, S)$ . In a context where you want the best of the available options, the threshold at  $w$  for  $\lceil S \text{ wants } p \rceil$ ,  $t_c^+(w, S)$ , is the expected value  $S$  at  $w$  assigns to the best of the available options, while for  $\lceil S' \text{ wants } p \rceil$ , the threshold at  $w$ ,  $t_c^+(w, S')$ , is the expected value  $S'$  at  $w$  assigns to the best of the available options. (More rigorously, it’s the option that the *agent finds best* among what *she takes* the available options to be.)

with Context-sensitive Threshold Profile is fit to capture this fact. In evaluating a want ascription at  $w_{@}$ , the threshold  $t_c(w_{@})$  equals the expected value Amelia at  $w_{@}$  assigns to watching the best available movie at  $w_{@}$ , *Hotel Transylvania*. At  $w_{Jaws}$ , the threshold is different—it's  $t_c(w_{Jaws})$ , the expected value Amelia at  $w_{Jaws}$  assigns to watching the best available movie at  $w_{Jaws}$ , *Jaws*.

The upgrade is needed only when ‘want’ is embedded. Conditionals, like (1-b), are one kind of embedding; attitude ascriptions, like ‘Otto thinks Esther wants to get pizza’, are another. Only when there’s a difference in the world of evaluation without a difference in context is there a difference between the predictions of the What’s-good-enough Account with Context-sensitive Threshold Profile and the un-upgraded What’s-good-enough Account with Context-sensitive Threshold. And this occurs only when ‘want’ is embedded.

## 8 On multiple-senses hypotheses

I have attributed the context-sensitivity of want ascriptions to *context-sensitivity in ‘want’*. As I mentioned in section 5, others attribute it to different senses of ‘want’—in other words, to *ambiguity in ‘want’*.

Such hypotheses are designed to make sense (with ambiguity) of certain purported phenomena involving want ascriptions—e.g. the purported phenomena that whatever you intend to do, you can be truly said to want. As I’ll argue, my account can make sense (without ambiguity) of these purported phenomena too. Further, my account is neutral on whether the phenomena are genuine, a desirable feature, since some have questioned the reality of these phenomena. I’ll illustrate all of this with two commonly discussed senses of ‘want’: *all-things-considered* ‘want’ and *volitive* ‘want’.

(Multiple-senses hypotheses should give us pause. According to a widely accepted methodological principle, we should prefer hypotheses on which a term is context-sensitive, like mine, to hypotheses on which it’s ambiguous.<sup>23</sup> And, as I showed in section 5, there aren’t exactly two senses of ‘want’, meaning that multiple-senses hypotheses can’t be ones on which ‘want’ is two-way ambiguous, like ‘bank’, but must be ones on which it is at least three-way ambiguous, an especially undesirable result.)

What you want, all things considered, is normally taken to be what you prefer

---

<sup>23</sup>Except in cases where languages lexicalize the term differently—e.g. how ‘bank’ can translate into German as either ‘Bank’ or ‘Ufer’.

most, given certain options. My account can make sense of such cases as follows. When we're in a context where what's at issue is what you prefer among most given certain alternatives, the threshold is set in a certain way. In particular, the threshold is equal to the expected value that you assign to the most preferred option, which is therefore wanted since its expected value equals, and so meets, the threshold. And none of the other options are wanted. Not being the most preferred option means a lower expected value than that of the most preferred, which is to say lower than the threshold.

The kind of context that some would say contains the all-things-considered 'want', then, just comes out as *a special case of a more general contextual variation in 'want'*. The threshold can be set in different places, and when it's set in a certain place—a place that, among the options you're confronted with, is met only by the most preferred option—it can match what others would say is a special sense of 'want'.

Further, we can stay neutral on a contentious thesis that surrounds all-things-considered wanting—the thesis that there's always a context where you can truly be said to want the most preferred of a given set of options, even when all of those options are repellent to you. Take the case where you must either shoot the one or let the two be killed. Is there a context where you're truly said to want to shoot the one, given that you most prefer it? The answer is yes according to the proponent of the thesis; for the denier, the answer is no. In terms of the threshold, the proponent will say that there's a context where the threshold is met by shooting the one; the denier will deny this.

More generally, the proponent will say that there is no *floor* on where the threshold can be set: there is always a context where the threshold is set to the expected value that you assign to the most preferred option, no matter how low that value is. The denier thinks that there is a floor. Both the proponent and the denier can accept my framework. And further, the framework gives us a clean way to state the point of their contention—in terms of whether the threshold has a floor.

Turn now to the controversial *volitive* sense of 'want'.<sup>24</sup> Volitive wanting is supposed to go with intending. If you intend *p*, then you volitively want *p*. Put without appealing to a special sense, the idea is that if you intend *p*, then there's a context where you're truly said to want *p*. We can accommodate this idea too with

---

<sup>24</sup>Davis (1984) reviews the literature.

context-sensitivity in ‘want’.<sup>25</sup>

In the terms of my view, the idea is that if you intend  $p$ , then there’s a context where the expected value you assign to  $p$  meets the threshold—a context where you’re truly said to want  $p$ . That idea and its denial are each compatible with my view, which says nothing on where the threshold can be set. And again, my view provides a clean framework to state the debate. This time the question is: does intending  $p$  entail the existence of a context where the threshold is met by the expected value you assign to  $p$ ?

Similar things will go, I hope, for other multiple-senses hypotheses, like *pro tanto* ‘want’, or Daveney’s (1961) *pro attitude* ‘want’, or Davis’s (1984) *appetitive* ‘want’. Supposing that our use of ‘want’ tracks such things as *pro tanto* wanting or *pro attitude* wanting or *appetitive* wanting—whatever these amount to—they can be accounted for with the threshold being set in some particular place. And if you’d like to deny the existence of any of these purported phenomena that the senses of ‘want’ are supposed to track, you can say that the threshold can’t be set in the relevant ways.

## 9 Conclusions

If folk psychology and descriptive decision theory can’t be understood in terms of one another, the dubious dualism looms: a competition between two greatly significant representations of our belief-like and desire- and preference-like states—two greatly significant systems of explanation and prediction of our actions.

To avoid the dubious dualism, I have taken up part of the project of accounting for the concepts of folk psychology—in particular, wanting—with the concepts of decision theory. I’ve maintained that you’re truly said to want what’s good enough in your eyes—to want that to which you assign a certain expected value—and what counts as good enough shifts by context.

This is just a first step. The next step, I believe, is investigating the logic of wanting. The logic of belief has proven instrumental in the debates about Lockeanism, both between Lockean and non-Lockean and among Lockean.<sup>26</sup> I’ll close by highlighting some issues in the logic of wanting to pursue further as we explore whether we can account for wanting in the terms of decision theory—and if so, how.

---

<sup>25</sup>Levinson (2003) and Condoravdi and Lauer (2016) propose to do the same (within different frameworks than mine).

<sup>26</sup>Thank you to an anonymous reviewer for pointing out the importance of the logic of wanting here.

Various debates about Lockeanism have been animated by whether belief is closed under multi-premise entailment (see e.g. (Foley, 1992), (Hawthorne and Bovens, 1999) and (Leitgeb, 2017, ch. 3)). Wanting, as my theory predicts, is *not* closed under multi-premise entailment: ‘I want to clean with bleach’ (it’ll kill germs) and ‘I want to clean with ammonia’ (it’ll remove grease) don’t entail ‘I want to clean with bleach *and* ammonia’ (together, they produce deadly fumes). It may well be, though, that some restricted form of multi-premise closure is valid. Some think that a restricted form of *single*-premise closure is valid for wanting,<sup>27</sup> even though unrestricted single-premise closure, which my theory doesn’t validate, is not: ‘I want to die quickly’ doesn’t entail ‘I want to die’ (Anand and Hacquard, 2013).

Questions about the logic of ‘ought’ also give clues for what to pursue in the logic of wanting. For example, Drucker (2017) points out that the Miner’s Puzzle (see e.g. (Kolodny and MacFarlane, 2010)) has an analogue with ‘want’, and Jerzak (2019) draws parallels between the so-called *objective* ‘ought’ and *subjective* ‘ought’,<sup>28</sup> and Jerzak’s *advisory* ‘want’ and *predictive* ‘want’, respectively.

Finally, the problematic inference pattern identified by Villalta (2008, p. 478) deserves our attention.<sup>29</sup> Crnić (2011) gestures at one way to invalidate the pattern—developed by Phillips-Brown (2018)<sup>30</sup>—and Rubinstein (2017) offers another. It remains to be seen whether these ways of blocking the inference pattern work within a decision-theoretic framework.

Even without having taken these next steps in the logic of wanting, our first step has given us a lot to work with. Building on the general project of context-sensitive understandings of attitude verbs, I’ve hypothesized a shifting bar for what counts as good enough—shifting with both context and world—a bar that explains why what you’re truly said to want is not intrinsic to you, why it varies by context as the communicative purposes of ascribers vary by context. Further, we can, if we want to, explain phenomena that have motivated traditional multiple-senses hypotheses, without committing to an unsavory ambiguity.

---

<sup>27</sup>Crnić (2011) reviews the literature.

<sup>28</sup>See e.g. (Carr, 2015) on the subjective ‘ought’.

<sup>29</sup>The inference pattern is this: if  $\ulcorner S \text{ wants } p \urcorner$  is true and  $S$  is certain that  $p$  iff  $q$ , then  $\ulcorner S \text{ wants } q \urcorner$  is true. (Villalta talks of belief, not certainty, but the point applies just as well to certainty.) Virtually every semantics for ‘want’, including mine, incorrectly validates this inference.

<sup>30</sup>Phillips-Brown’s question-based semantics resembles Dandele’s (ms) situation semantics.

## References

- Pranav Anand and Valentine Hacquard. Epistemics and attitudes. *Semantics and Pragmatics*, 6(8):1–59, 2013.
- Nomy Arpaly and Timothy Schroeder. *In Praise of Desire*. Oxford University Press, 2013.
- Lauren Ashwell. Conflicts of desire: Dispositions and the metaphysics of mind. In Jonathan Jacobs, editor, *Causal Powers*. Oxford University Press, 2017.
- Derek Baker. Ambivalent desires and the problem with reduction. *Philosophical Studies*, 150(1):37–47, 2010.
- Ken Binmore. Interpersonal comparison of utility. In Don Ross and Harold Kincaid, editors, *The Oxford Handbook of Philosophy of Economics*. Oxford University Press, 2009.
- Kyle Blumberg and Ben Holguín. Embedded Attitudes. *Journal of Semantics*, March 2019.
- Richard Bradley. Conditional desirability. *Theory and Decision*, 47(1):23–55, 1999.
- Lara Buchak. *Risk and Rationality*. Oxford University Press, 2013.
- Lara Buchak. Belief, credence, and norms. *Philosophical Studies*, 169(2):1–27, 2014.
- Jennifer Carr. Subjective ought. *Ergo*, 2(27):678–710, 2015.
- Cleo Condoravdi and Sven Lauer. Anankastic conditionals are just conditionals. *Semantics and Pragmatics*, 9(8):1–69, November 2016.
- Luka Crnić. *Getting even*. PhD thesis, Massachusetts Institute of Technology, 2011.
- Sophie Dandeleit. Partial desires, blinkered beliefs. Unpublished manuscript, ms.
- T. F. Daveney. Wanting. *Philosophical Quarterly*, 11(April):135–144, 1961.
- Donald Davidson. Actions, reasons, and causes. *Journal of Philosophy*, 60(23):685–700, 1963.
- Wayne Davis. The two senses of desire. *Philosophical Studies*, 45(2):181–195, 1984.
- James Dreier. Pettit on preference for prospects and properties—discussion. *Philosophical Studies*, 124(2):199–219, 2005.
- Fred Dretske. *Explaining Behavior: Reasons in a World of Causes*. MIT Press, 1988.
- Daniel Drucker. Policy externalism. *Philosophy and Phenomenological Research*, 94(3), 2017.
- Kenny Easwaran. Dr. Truthlove or: How I learned to stop worrying and love Bayesian probabilities. *Noûs*, 50(4):816–853, 2016.
- P. Erickson, J.L. Klein, L. Daston, R. Lemov, T. Sturm, and M.D. Gordin. *How Reason Almost Lost Its Mind: The Strange Career of Cold War Rationality*. University of Chicago Press, 2013.
- Kai von Fintel. Npi licensing, strawson entailment, and context dependency. *Journal of Semantics*, 16(2):97–148, 1999.
- Richard Foley. *Working Without a Net: A Study of Egocentric Epistemology*. Oxford University Press, 1992.
- Dorit Ganson. Evidentialism and pragmatic constraints on outright belief. *Philosophical Studies*, 139(3):441–458, 2008.
- James Hawthorne and Luc Bovens. The preface, the lottery, and the logic of belief. *Mind*, 108(430):241–264, 1999. doi: 10.1093/mind/108.430.241.
- Irene Heim. Presupposition projection and the semantics of attitude verbs. *Journal of Semantics*, 9(3):183–221, 1992.
- Frank Jackson. Internal conflicts in desires and morals. *American Philosophical Quarterly*, 22(2):105–114, 1985.

- Ethan Jerzak. Two ways to want? *Journal of Philosophy*, 116(2):65–98, 2019.
- Niko Kolodny and John MacFarlane. Ifs and oughts. *Journal of Philosophy*, 107(3):115–143, 2010. doi: 10.5840/jphil2010107310.
- Daniel Lassiter. *Measurement and Modality: the Scalar Basis of Modal Semantics*. PhD thesis, New York University, 2011.
- Hannes Leitgeb. *The Stability of Belief: How Rational Belief Coheres with Probability*. Oxford University Press, 2017.
- Dmitry Levinson. Probabilistic model-theoretic semantics for *want*. *Semantics and Linguistics Theory (SALT)* 13, pages 222–239, 2003.
- David Lewis. *On the Plurality of Worlds*. Blackwell, 1986.
- Dustin Troy Locke. The decision-theoretic lockean thesis. *Inquiry: An Interdisciplinary Journal of Philosophy*, 57(1):28–54, 2014.
- Michelle Montague. Against propositionalism. *Noûs*, 41(3):503–518, 2007.
- Carolyn R. Morillo. The reward event and motivation. *Journal of Philosophy*, 87(4):169–186, 1990.
- Philip Pettit. Decision theory and folk psychology. In Michael Bacharach and Susan Hurley, editors, *Essays in the Foundations of Decision Theory*, pages 147–175. Blackwell, 1991.
- Philip Pettit. *Rules, Reasons, and Norms: Selected Essays*. Clarendon Press, 2002.
- Milo Phillips-Brown. I want to, but... In Robert Truswell, Chris Cummins, Caroline Heycock, Brian Rabern, and Hannah Rohde, editors, *Sinn und Bedeutung 21*, pages 951–968. 2018.
- John L. Pollock. *Thinking About Acting: Logical Foundations for Rational Decision Making*. Oxford University Press, 2006.
- Robert van Rooij. Some analyses of pro-attitudes. In Henriette de Swart, editor, *Logic, Game Theory and Social Choice*, pages 534–548. Tilburg University Press, 1999.
- Aynat Rubinstein. Straddling the line between attitude verbs and necessity modals. In *Modality Across Syntactic Categories*, pages 610–633. Oxford University Press, 2017.
- Michael Smith. *The Moral Problem*. Blackwell, 1994.
- Julia Staffel. Beliefs, buses and lotteries: Why rational belief can't be stably high credence. *Philosophical Studies*, 173(7):1721–1734, 2016.
- Robert Stalnaker. A theory of conditionals. In Nicholas Rescher, editor, *Studies in Logical Theory (American Philosophical Quarterly Monographs 2)*, pages 98–112. Blackwell, 1968.
- Robert Stalnaker. *Inquiry*. Cambridge University Press, 1984.
- Robert Stalnaker. *Our Knowledge of the Internal World*. Oxford University Press, 2008.
- Katie Steele and H. Orri Stefánsson. Decision theory. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter edition, 2016.
- G. C. Stine. Skepticism, relevant alternatives, and deductive closure. *Philosophical Studies*, 29(4): 249–261, 1976.
- Scott Sturgeon. Reason and the grain of belief. *Noûs*, 42(1):139–165, 2008.
- Elisabeth Villalta. Mood and gradability: An investigation of the subjunctive mood in Spanish. *Linguistics and Philosophy*, 31(4):467–522, 2008.
- Chase Wrenn. A puzzle about desire. *Erkenntnis*, 73(2):185–209, 2010.
- Seth Yalcin. Belief as question-sensitive. *Philosophy and Phenomenological Research*, 93(2), 2016.