Desiderative Lockeanism[1]

this is a draft; the final version is forthcoming in *Australasian Journal of Philosophy*

Milo Phillips-Brown, University of Edinburgh

Abstract. According to the *Desiderative Lockean Thesis*, there are necessary and sufficient conditions, stated in the terms of decision theory, for when one is truly said to want. I advance a new Desiderative Lockean view. My view is distinctive in being *doubly context-sensitive*. What a person is truly said to want varies by context, a fact that others attempt to capture by positing a single context-sensitive parameter to evaluate want ascriptions; I posit two. Only with a doubly context-sensitive view can we explain a range of facts that go unexplained by all other Desiderative Lockean views.

Keywords: desire, desire ascriptions, decision theory, context-sensitivity, Lockean Thesis

1 Introduction

The Lockean Thesis is the view that there are necessary and sufficient conditions, stated in the terms of decision theory—specifically, credences—for when one is truly said to believe.[2] An analogous thesis for wanting says that there are necessary and sufficient conditions, stated in the terms of decision theory, for when one is truly said to want.[3] Call this the *Desiderative Lockean Thesis*.

Together, the Lockean and Desiderative Lockean Theses unify two enormously significant systems of how the mind relates to action (Phillips-Brown, 2021). In daily life, in philosophy, and in psychology, we employ the so-called belief–desire model of the mind—i.e. a model of believing and wanting—to predict and explain action (e.g. Davidson (1963)). In the quantitative social sciences and philosophy, we likewise use decision theory's model of the mind—a model of credence and preference—to predict and explain action (e.g. Bermúdez (2009)).

I propose a new Desiderative Lockean view that is primarily motivated by the context-sensitivity of want ascriptions. It is well established that we can hold fixed an agent's mental states and yet what she can be truly said to want differs by context: she can be truly said to want

[1]For feedback on drafts and countless conversations, thank you to Kyle Blumberg, Marion Boulicault, Carol Brown, Nilanjan Das, Kai von Fintel, Lyndal Grant, Justin Khoo, Matt Mandelkern, Ginger Schultheis, Jen Semler, Kieran Setiya, Jack Spencer, Quinn White, Steve Yablo, two anonymous reviewers, and editors at this journal. Thank you especially to David Boylan for years of help and guidance.

[2]See e.g. Foley (1992) and Easwaran (2016).

[3]The Desiderative Lockean views so far proposed come from van Rooij (1999), Levinson (2003), Wrenn (2010), Lassiter (2011), Drucker (2017), Jerzak (2019), Phillips-Brown (2021), and Blumberg (2023).

something in one context but not in another.[4] The extent of this context-sensitivity is even greater than has been appreciated: I supply new sorts of examples in this paper. The extent of this context-sensitivity also calls for a more contextually flexible semantics than Desiderative Lockeans have appreciated. Whereas other Desiderative Lockeans have posited one context-sensitive parameter, I posit two—my view is *doubly context-sensitive*. Only with this greater flexibility can we explain a range of phenomena—including both context-sensitivity and the closure behavior of desire ascriptions—that go unexplained by extant Desiderative Lockean views.

2 Basics of Desiderative Lockeanism

According to Desiderative Lockeanism, there are necessary and sufficient conditions, stated in the terms of decision theory, for when one is truly said to want. Which terms of decision theory? The common answer is *expected value*, which I will understand in roughly the Jeffreyan (1965) tradition. Given the set of possible worlds W, an agent S, a credence function Cr_S, and a utility function u_S, the expected value S assigns to $p =_{\text{df}} \sum_{w_i \in W} Cr_S(w_i \mid p) \cdot u_S(w_i)$. Some caution that expected value may not ultimately be the right decision-theoretic entity to account for wanting.[5] Perhaps the best Desiderative Lockean view will be stated in terms of expected value; perhaps it will be stated in other decision-theoretic terms. For simplicity, I'll work with expected value.

Desiderative Lockean views are varied. Phillips-Brown (2021) taxonomizes them into two species. On *what's-good-enough* accounts—which I favor—one wants what's good enough in her eyes, relative to some threshold of desirability (Lassiter, 2011; Phillips-Brown, 2021). On *what's-best accounts*—which I argue against just below—one wants something just if she most prefers it to certain alternatives: just if it's best when compared to those alternatives (Levinson, 2003; Jerzak, 2019). For concreteness, consider simple versions of the two kinds of views, both entertained by van Rooij (1999) and the latter endorsed by Levinson (2003).

> **Simple What's-good-enough Account**[6]
> ⌜S wants p⌝ is true iff the expected value S assigns to p meets a certain threshold.

> **Simple What's-best Account**
> ⌜S wants p⌝ is true iff the expected value S assigns to p exceeds the expected values she assigns to certain alternatives.

To see the flaw in what's-best accounts, imagine that you are deciding where to go to dinner (Phillips-Brown, 2021). There are three options: the pizzeria, the ramen shop, and the hot dog stand. Pizza and ramen both sound good; hot dogs don't; pizza is most appealing. You could say:

[4] See e.g. Davis (1984), Villalta (2008), Lassiter (2011), and Phillips-Brown (2021).

[5] For example, Levinson (2003) and Wrenn (2010).

[6] 'S' ranges over names of agents; 'S' ranges overs the corresponding agents; 'p' ranges over terms that denote propositions; 'p' ranges over the corresponding propositions.

(1) I want to go to the ramen shop.

(2) And I want to go to the pizzeria even more.

(1) is true, but going to the ramen shop is not best, falsifying the Simple What's-best Account (you assign a higher expected value to the ramen shop than to the pizzeria) and what's-best accounts generally. The Simple What's-good-enough Account gets the case right: the pizzeria and the ramen shop can each be good enough; the expected values you assign to each can meet the threshold.

The Simple What's-good-enough Account needs improving, though, since it does not admit of context-sensitivity, and as I noted in §1, want ascriptions are context-sensitive. Consider a case. Johnnie must do one of two chores: mow the lawn or wash the dishes, both of which he hates. He prefers mowing the lawn to washing the dishes, though, and so that is what he'll do. He's now walking to the shed, where the lawn mower is stored:

Conversation One

(3) I see that Johnnie is going to mow the lawn. Is that something he wants to do?

(4) No, Johnnie doesn't want to mow the lawn (he's only going to because he has to do a chore and mowing the lawn is better than washing the dishes).

Conversation Two

(5) Why is Johnnie going to the shed?

(6) Johnnie wants to mow the lawn (and that's where the mower is—of course he hates mowing the lawn, but it's better than washing the dishes).

Viewed through the lens of what's-good-enough accounts, it's natural to see the context shift in the truth value of (6) as a context shift in *what counts as good enough* (Phillips-Brown, 2021). In the context where (6) is true, mowing the lawn is good enough, but it's not good enough in the context where it is false (i.e. in the context where (4) is true). Phillips-Brown regiments this intuition by allowing contextual variation in the threshold. The result is what I take to be the leading Desiderative Lockean view:

Context-sensitive-threshold Account (Phillips-Brown, 2021)
⌜S wants p⌝ is true in c iff the expected value S assigns to p meets the threshold in c.

The expected value that Johnnie assigns to mowing the lawn meets the threshold in the context where (6) is true, but not in the context where (6) is false.

3 Doubly context-sensitive Desiderative Lockeanism

3.1 The account

By adopting a what's-good-enough account with a context-sensitive threshold of desirability, Desiderative Lockeans can explain certain forms of context-sensitivity in want ascriptions. Yet there is a range of other phenomena—which I present later in §3—that go unexplained by a what's-good-enough account with a context-sensitive threshold (and indeed all other Desiderative Lockean views). Among these phenomena are cases of context-sensitivity in want ascriptions that cannot be traced to a context-sensitive threshold. Rather, I argue, these cases must be understood in terms of a different context-sensitive parameter. On the view I propose, want ascriptions are evaluated against a context-sensitive threshold *and* a second context-sensitive parameter, a set of propositions that I call a *prospect set*. In being *doubly context-sensitive* my view stands apart from Desiderative Lockean views. I'll set out my account here in §3.1 and then in the rest of §3, discuss the phenomena that motivate it.

Let's begin with terminology. A *prospect*, as I understand it, is a proposition, and a prospect set is, well, a set of prospects. We'll say that a prospect p_i that entails p—i.e. a proposition p_i that entails p—is a *p-prospect*. You can think of a p-prospect as a *way* for p to come about.[7] For example, the proposition that Andres pursues a PhD in anthropology is an *Andres-pursues-a-PhD* prospect; it is a way for Andres to pursue a PhD. Similarly, a prospect that entails $\neg p$—which I'll call a *¬p-prospect*—is a way for $\neg p$ to come about. Or, put another way, it is an *alternative* to p. Pursuing an MD (instead of a PhD) is a way of not pursuing a PhD; it is an alternative to pursuing a PhD.

According to the what's-good-enough accounts so far proposed, you're truly said to want p in c just if p is good enough in your eyes. According to my what's-good-enough account, you're truly said to want p just if a *p-prospect* is good enough in your eyes. ⌜S wants p⌝ is evaluated against both a context-sensitive threshold and a context-sensitive prospect set. ⌜S wants p⌝ is true in c just if there is a p-prospect—*within the prospect set determined by c*—that is good enough relative to the threshold in c. More precisely still:

> **Doubly Context-sensitive Account** (my account)
> ⌜S wants p⌝ is true in c iff there is a p-prospect p_i in the prospect set in c such that: the expected value S assigns to p_i meets the threshold in c.

At this point, the view may feel somewhat abstract. It will become more concrete as you see it in action throughout the rest of §3.

[7] In using 'ways' like this, I follow (Cariani, 2013) in the context of deontic modals and (Grant and Phillips-Brown, 2020) in the context of wanting.

3.2 Context-sensitivity in *p*-prospects

Consider a case:

> Mo is a competitive runner who gets a special thrill from winning races. He does not, however, value winning *per se*, but rather winning *fairly*; winning by cheating is worse than not winning at all. Mo has registered for an upcoming race, but he knows that faster runners than him have also registered; he can win only if he cheats. He has therefore resolved not to train very hard.

We can imagine two different conversations that might go on:

Conversation Three

(7) I heard Mo's not going to train very hard for this race. Doesn't he want to win?

(8) Of course Mo wants to win this race. (But he can only win by cheating, which is why he's not going to train very hard.)

Conversation Four

(7) I heard Mo's not going to train very hard for this race. Doesn't he want to win?

(9) No, he doesn't want to win this race. (He can only win by cheating, which is why he's not going to train very hard.)

What Mo is truly said to want differs by context: in the context where (8) is true, he's truly said to want to win, but in the context where (9) is true, he's not. And yet Mo's mental states are fixed across contexts: he likes winning fairly, he dislikes winning by cheating, and he knows that he can only win the upcoming race by cheating. Indeed, we may imagine that (8) and (9) are asserted at the same time.

A context-sensitive threshold cannot explain Mo's case. To see why, consider how the Context-sensitive-threshold Account (from §2) treats the case. That view tells us that ⌜S wants p⌝ is true in *c* just if *p* is good enough for *S* by the standard in *c*—i.e. the expected value *S* assigns to *p* meets the threshold in *c*. Mo expects that if he wins, he will win by cheating, and so the expected value that he assigns to winning equals (or more or less equals) the expected value he assigns to winning by cheating. The account would tell us then, that (8) is true in *c* just if winning by cheating is good enough for Mo in *c*. The context-sensitivity in (8), we'd be led to believe, is due to winning by cheating being good enough in one context but not in another.

This is not right. First, because in order for (8) to be true, there would have to be a context where winning by cheating is good enough, and so a context where Mo would be truly said to want to win by cheating—but Mo cannot be truly said to want to win by cheating.

Second, because what the account tells us about the case doesn't comport with its intuitive understanding. The account tells us that there is a single way of Mo winning—winning by cheating—that is good enough in one context but not in another. Intuitively, though, *different ways of Mo winning* are in some sense being talked about in the different contexts. In contexts where it's true to say that Mo wants to win, it's true because we are somehow talking about him winning fairly, which he looks upon favorably. In contexts where it's false, it's false because we are somehow talking about him winning by cheating, which he looks upon disfavorably.

This intuitive understanding is neatly captured by my view. Different *Mo-wins* prospects represent different ways for Mo to win; a difference in which ways are being talked about in a context amounts to a difference in which *Mo-wins* prospects appear in the prospect set. Mo is truly said to want to win in a context whose space contains the *Mo-wins* prospect that Mo wins fairly, which is good enough (indeed, it's great).[8] Mo is not truly said to want to win in a context whose prospect set lacks this *Mo-wins* prospect and instead contains the *Mo-wins* prospect that Mo wins by cheating—in such a context, there is no *Mo-wins* prospect that is good enough.

3.3 Context-sensitivity in $\neg p$-prospects

Consider another case:[9]

> Li is the Secretary of State, and he prefers diplomacy to violence. The country faces two enemies: the pretty bad guys and the very bad guys. Li had proposed to the President three options, which in descending order of Li's preference are these: negotiate with the very bad guys, bomb the very bad guys, bomb the pretty bad guys. (Best to bomb no one, but if someone is to be bombed, it should be the very bad guys, not the pretty bad guys.) The President immediately dismisses the idea of negotiating with the very bad guys and says she'll soon decide between the remaining two options.

Imagine utterances of the following sentences, where all caps indicates emphasis:

(10) Li wants the President to bomb the VERY BAD GUYS.

(11) Li wants the President to BOMB the very bad guys.

It would be natural to hear an utterance of (10) as true and an utterance of (11) as false. Whether Li is truly said to want the President to bomb the very bad guys varies by context, even though—we

[8] There is a complication here. If Mo assigns credence zero to the proposition that he wins fairly, then certain understandings of expected value (including the one I adopt) dictate that the expected value Mo assigns to winning fairly is undefined (Wrenn, 2010). If it's undefined, then it does not meet the threshold, of course. If it doesn't meet the threshold, then my view would incorrectly predict that it's not good enough for Mo. This problem is not unique to my view; every existing Desiderative Lockean that uses expected value runs into a similar problem (Wrenn, 2010). Because the problem is not special to me, I won't try to resolve it; see (Wrenn, 2010) and (Blumberg, 2023) for promising solutions.

[9] The structure of the case comes from (Villalta, 2008) and its details come from (Grano and Phillips-Brown, 2022).

may imagine—his mental states are fixed. (His beliefs about the situation may remain the same, as may his preferences between the three options he presented to the President.) Indeed, we may imagine—as we have in previous cases of context sensitivity—that utterances of (10) and (11) occur at the same time (within different conversations).

Why the difference in what Li is truly said to want? Intuitively, because bombing the very bad guys is being compared to different alternatives. (10) evokes one alternative to bombing the very bad guys—bombing the pretty bad guys instead—and in contrast to this alternative, Li is truly said to want the President to bomb the very bad guys. But Li is not truly said to want the President to bomb the very bad guys when doing so is contrasted with a different alternative—the alternative, evoked in (11), of instead negotiating with the very bad guys.

My view is ready-made to regiment the intuition that the difference in contexts owes to a difference in alternatives to p. Contexts may differ in their prospect sets, and as noted above, certain prospects represent alternatives to p—namely, $\neg p$ prospects.[10] Contrast, for example, the following two prospect sets that differ in their alternatives to bombing the very bad guys. Where 'VBGs' is short for 'very bad guys' and 'PBGs' is short for 'pretty bad guys':

prospect set the where alternative to bombing VBGs is bombing PBGs	
bomb very bad guys	bomb pretty bad guys

prospect set where the alternative to bombing VBGs is negotiating with them	
bomb very bad guys	negotiate with very bad guys

Imagine, for concreteness, that in this bombing case, the speakers are interested in *all-things-considered wanting*. That is, they are in a context where the agent is truly said only to want the best of the available options. In the terms of my view, this is a context c in which the threshold is equal to the expected value that the agent assigns to their most preferred prospect in c—that is, the prospect to which the agent assigns the highest expected value. So, the threshold in c for our bombing cases equals the expected value that Li assigns to the best prospect in c—the best prospect is wanted. My view then rightly predicts that Li is truly said to want to bomb the very bad guys in the context whose prospect set is the first of those just above: bombing the very bad guys is the best prospect when the alternative is bombing the pretty good guys. The view also rightly predicts that Li isn't truly said to want to bomb the very bad guys in the context whose prospect set is the second of those just above: bombing the very bad guys isn't the best prospect when the alternative is negotiating with them.

[10] I am taking inspiration from Villalta (2008) and Lassiter (2011).

3.4 Single-premise closure and context-sensitivity in the prospect set

Existing Desiderative Lockean theories falter with the cases from §3.2 and §3.3. To get right what these theories get wrong, we need a context-sensitive prospect set, in addition to a context-sensitive threshold. This much we've seen. There is more still. In having two context-sensitive parameters, my view allows us to answer a hotly debated question: does wanting obey *single-premise closure*?[11] (Single-premise closure is also known as *upward monotonicity*.) Put roughly, the question is: if ⌜S wants p⌝, and *p* entails *q*, can one infer that ⌜S wants q⌝ is true?

In opposition to other Desiderative Lockeans, I answer yes—but a restricted yes. As I explain below, I follow von Fintel (1999) in saying that single-premise closure is what's known as *Strawson valid*. (Put another way: want ascriptions are *Strawson upward monotonic*.) §3.4.1–§3.4.3 concern different sets of data relating to single-premise closure, as well as further cases in which what the agent can be truly said to want varies by context, even holding her mental states fixed. Previous Desiderative Lockean views explain none of the data; my view explains them by appeal to context-sensitivity of the prospect set.

3.4.1 Single-premise closure: round one

Inferences of single-premise closure look impeccable in many cases, as others have pointed out.[12] Where $\rightsquigarrow$ indicates felt entailment:

(12) a. Andres wants to pursue a PhD in anthropology.
b. $\rightsquigarrow$ Andres wants to pursue a PhD.

(13) a. James wants to eat his dessert and relish every bite.
b. $\rightsquigarrow$ James wants to eat his dessert.

And the inference blatantly fails in other cases. Imagine, for example, a scenario adapted from Crnič (2011), who draws from (Prior, 1958). Jane believes that Cy will be robbed and that it's too late to stop the robbery:

(14) a. Jane wants to help Cy after he's robbed.
b. $\not\rightsquigarrow$ Jane wants Cy to be robbed.

Or take a case adapted from (Anand and Hacquard, 2013):

(15) a. Pete wants to die peacefully.
b. $\not\rightsquigarrow$ Pete wants to die.

This is our first set of data surrounding single premise closure: it's strongly felt to hold in certain cases, like in (12) and (13), but strongly felt not to hold in others, such as (14) and (15).

[11] See e.g. (Heim, 1992; von Fintel, 1999).

[12] See e.g. (Crnič, 2011).

On my view, it's easy to predict cases where single-premise closure is strongly felt to hold. Suppose that ⌜S wants p⌝ is true in context c. Then, my view tells us, there is a p-prospect p_i in the prospect set in c that is good enough for S by the standards in c. If p entails q, then p_i is thereby a q-prospect, and one that's good enough in c—from which it follows that ⌜S wants q⌝ is true in c. The inferences in (12) and (13) go through.

It takes a subtler hand to explain cases where single-premise closure is strongly felt to fail. I hypothesize that in such cases, the premise and the conclusion are evaluated in different contexts. So, for example, in the context where (14-a) is most naturally evaluated, it is true; but (14-b) is most naturally evaluated in a different context, one where it is false.

This hypothesis may strike you as familiar, for it is analogous to the one made by many about Prior's (1958) Good Samaritan Paradox. Imagine, as Prior has us do, that Cy will be robbed:

The Good Samaritan Paradox

(16) a. It ought to be the case that Jane helps Cy after he's robbed.

b. $\nRightarrow$ It ought to be the case that Cy is robbed.

The Good Samaritan Paradox is called a paradox because the inference from (16-a) to (16-b) is fallacious and yet it is validated in what are known as standard deontic logics for 'ought'—logics on which 'ought' validates single-premise closure (Navarro and Rodríguez, 2014). (In the literature on deontic logic and modals, single-premise closure sometimes goes by the name *inheritance*.) These logics have fallen by the wayside, but more sophisticated accounts of 'ought', like Kratzer's (1991), still validate single-premise closure in one form or other. The challenge for the Kratzerian, then—like the challenge for me—is to explain cases of apparent failures of single-premise closure, like those in the Good Samaritan Paradox. What Kratzerians do is maintain—as I have—that such cases exhibit context shift. Kratzerians say, for example, that in the context where (16-a) is most naturally evaluated, it is true; but (16-b) is most naturally evaluated in a different context, one where it is false (von Fintel and Heim, 2011).

Why think that (16-a) and (16-b) are evaluated in different contexts? The basic idea is that when we're evaluating in a context how the world ought to be, we take certain things *as given* (von Fintel and Heim, 2011). When we assess how the world ought to be, full stop—when we assess what the ideal world looks like—it ought not to be that Jane helps Cy after he's robbed; in the ideal world, Cy is not robbed at all. Rather, it ought to be that Jane helps Cy after he's robbed only *given that he will be robbed*. We take it as given that Cy will be robbed when evaluating (16-a), which is why the sentence is true. In evaluating (16-b) we rather evaluate what ought to be, full stop. (One clue to this difference in what's taken as given is that (16-a) presupposes that Cy will be robbed, but (16-b) does not.)

This notion of taken as given is also at play in want ascriptions—or so I hypothesize. When we're evaluating in a context what an agent wants, we take certain things as given. Does Jane want, full stop, for the world to be such that Cy is robbed and she helps him after? No! She only wants this to be the case *given that Cy will be robbed*. (14-a), 'Jane wants to help Cy after he's robbed', is most naturally evaluated in a context where this is indeed taken as given. Not so in the context where (14-b), 'Jane wants Cy to be robbed', is most naturally evaluated. (One clue to this difference in what's taken as given is that (14-a) presupposes that Jane *believes* that Cy will be robbed, but (14-b) does not—for, as Karrttunen (1973) first observed, ⌜S wants p⌝ presupposes that *S* believes the presuppositions of *p*.)

So, we have what we were after: a story of why (14-a) and (14-b) are evaluated in different contexts—a story that explains apparent failures of single-premise closure. Let us now spell it out more precisely in the terms of my view.

First, we make precise the notion of taken as given—and in this we'll again follow Kratzer (she doesn't use the phrase 'taken as given', but the idea is there). The exact details of her view aren't important to see my point; what's important to know is that an 'ought' sentence is evaluated against a contextually determined information state—a set of worlds known as the *modal base*. A proposition *p* is taken as given in a context *c* just if *p* is true at every world in the modal base in *c*. For me, a want ascription is analogously evaluated against a contextually determined information state—a set of propositions, the prospect set. To extend the analogy, we'll say that *p* is taken as given in *c* just if *p* is entailed by every prospect in the prospect set in *c*.

Second, we'll introduce a *diversity constraint*—which we find in Rubinstein's (2017) semantics for 'want'. The constraint ensures that ⌜S wants p⌝ is not evaluated in a context in which *p* is taken as given.

Diversity Constraint

⌜S wants p⌝ has a truth value in *c* only if:

(i) The prospect set in *c* contains *p*-prospects;

(ii) The prospect set in *c* contains $\neg p$-prospects.

Clause (i) ensures that want ascriptions are not vacuously false, for reasons I explain in footnote 13.[13] It is clause (ii) that corresponds to the idea that ⌜S wants p⌝ should not be evaluated in a context where *p* is taken as given: *p* is taken as given in *c* just if every proposition in the prospect set in *c* entails *p*, which is to say that *c* doesn't contain any $\neg p$-prospects—a violation of clause (ii). We've said that, intuitively, (14-b) shouldn't be evaluated in a context where it's taken as given that Cy will be robbed. This intuition is now regimented: the Diversity Constraint says that the sentence lacks a truth value in such a context.

The third thing to establish is that my view predicts not only the failure of the inference of (14-a) to (14-b), but also that (14-a) is true in the context where it's evaluated and that (14-b) is false

[13]If clause (i) is violated, then there aren't any *p*-prospects in the prospect set in *c*. *A fortiori* there aren't any *p*-prospects in that prospect set that are good enough for *S*, falsifying ⌜S wants p⌝.

in the context where *it's* evaluated. For concreteness, let's suppose that in these contexts, we're interested in all-things-considered wanting. In evaluating (14-a) in a context, we take as given that Cy will be robbed: every proposition in the prospect set entails that he will be robbed. What is this prospect set? The following is as natural candidate as any:

prospect set where it's taken as given that Cy will be robbed	
Cy will be robbed & Jane helps	Cy will be robbed & Jane doesn't help

The best prospect here is *Cy will be robbed and Jane helps*. This means, on my view, that (14-a) is true.

In evaluating (14-b) in a context, we do not take as given that Cy will be robbed: some prospects in the prospect set entail that he won't be. What is this prospect set? Something like this:

prospect set where it's not taken as given that Cy will be robbed	
Cy will be robbed	Cy won't be robbed

The best prospect here is *Cy won't be robbed*; my view predicts that (14-b) is false.

The fourth and final thing is that we are now in a position to see what I noted earlier: that my view, following von Fintel's, validates Strawson single-premise closure, which I state below. The core notion of Strawson validity in general (as opposed to just the Strawson validity of single-premise closure) is that a sentence may lack a truth value in a given context. Understood one way, the idea is that when a sentence lacks a truth value in a context, it is most naturally evaluated in another context. (Some sentences may lack a truth value in every context, but such sentences are not our concern here.) We've said, for example, that ⌜S wants p⌝ is most naturally evaluated in a context where *p* is not taken as given, and that ⌜S wants p⌝ lacks a truth value in such a context (by the Diversity Constraint). Roughly, Strawson single-premise closure says that in a context where both the premise and the conclusion have truth values, the conclusion is true if the premise is. Or more precisely:

Strawson single-premise closure

a. ⌜S wants p⌝ is true in *c*.
b. *p* entails *q*.
c. ⌜S wants q⌝ is either true or false in *c*.
d. $\Rightarrow$ ⌜S wants q⌝ is true in *c*.

The reason my view validates Strawson single-premise closure will be familiar. It's the same the reason, which I articulated on page 9, of why my view predicts cases where single-premise closure is felt to hold. Suppose that p entails q and that ⌜S wants p⌝ and ⌜S wants q⌝ both have truth values in c. If ⌜S wants p⌝ is true in c, then there is a p-prospect p_i in the prospect set in c that's good enough by the standards in c; that p_i is a q-prospect (because p entails q), and one that's good enough by the standards in c: ⌜S wants q⌝ is true in c.

Strawson single-premise closure enables us to precisely characterize the apparent failures of single-premise closure we find in (14-a) and (14-b). Contrast Strawson single-premise closure with:

Weak Single-premise Closure

a. ⌜S wants p⌝ is true in c.
b. p entails q.
c. $\Rightarrow$ ⌜S wants q⌝ is true in some context c'.

(14-a) and (14-b) *do* furnish a counterexample to weak single-premise closure: while (14-a) is true in some contexts, (14-b) is true in no context. But they do not furnish a counterexample to Strawson single-premise closure. The context where (14-a) is true is not one where (14-b) has a truth value: when we evaluate (14-a), it's taken as given that Cy will be robbed; (14-b) lacks a truth value in such contexts. So there is no context in which (14-a) is true, both (14-a) and (14-b) have truth values, and (14-b) is true. Strawson single-premise closure is intact.[14]

3.4.2 Single-premise closure: round two

Take a case, slightly modified from (Asher, 1987), that some have thought threatens single-premise closure. Andy is in New York, and he wants to get to London as quickly as possible. The best means to that end is to take the Concorde, but—as Andy well knows—to take a flight on the Concorde, one must pay, and the price of a ticket is prohibitively high for Andy. Consider:

(17) Andy doesn't want to take a flight on the Concorde. (It's too expensive.)

(18) Andy wants to take a free flight on the Concorde. (He knows that won't happen, though.)

We can judge both (17) and (18) as true, a pair of judgments that seem to result in a failure of single-premise closure.

Just what to make of these judgments is complicated when we consider their conjunction (von Fintel, 1999):

(19) ?? Andy wants to take a free flight on the Concorde and Andy doesn't want to take a flight on the Concorde.

The challenge here is to explain how (17) and (18) can both be judged true, and yet the conjunction, (19), is somehow defective—it is what we might call an *uncomfortable conjunction*. (An

[14]My hypotheses about apparent failures of single-premise closure are inspired by von Fintel (1999) and Crnič (2011).

uncomfortable conjunction is like what others call an *abominable conjunction*, just not as bad.)

When we judge (17) as true, why—intuitively speaking—do we judge it as true? Because, I take it, Andy has a negative attitude towards taking a *paid* trip on the Concorde: the badness of paying outweighs the goodness of getting to London quickly. Andy's attitude towards taking a paid trip on the Concorde matters, since in deciding whether to take a trip on the Concorde, Andy must evaluate not just the speed with which he'd get to London on the Concorde, but also its price. Andy's *decision problem* is between taking a paid trip on the Concorde and not taking a trip on the Concorde at all. We can represent a decision problem within my view with the prospect set: each decision Andy could make is a prospect.

prospect set that represents Andy's decision problem	
paid trip on Concorde	no trip on Concorde

In a context where this is the prospect set, (17) is true on my view. This prospect set contains only one *Andy-takes-a-flight-on-the-Concorde* prospect, and it is not a prospect that is good enough in Andy's eyes—for Andy, taking a paid trip on the Concorde is bad.

Turn to (18). Why, intuitively, is *it* judged true? Because Andy would enjoy a free trip on the Concorde! Because, in the terms of my view, a free trip on the Concorde is good enough for Andy; in the context where (18) is true, there is a *free-trip*-prospect that's good enough. This can't be the prospect set that represents Andy's decision problem, for in that prospect set, there are no *free-trip-on-the-Concorde* prospects at all. (Taking a free trip on the Concorde is, after all, not part of Andy's decision problem: taking a free trip on the Concorde is not something that he can decide to do.) Rather, it must be a prospect set like the following:

prospect set that includes a free-trip-on-the-Concorde prospect		
paid trip on Concorde	no trip on Concorde	free trip on Concorde

My view predicts that in a context where this is the prospect set, (18) is true. There is a *free-trip-on-the-Concorde* prospect that is not just good enough but great; getting to London as quickly as is possible without paying is an ideal outcome for Andy.

(17) and (18) are true relative to different prospects spaces. They are true, my view predicts, in different contexts. Why then is their conjunction, (19), uncomfortable? Because wanting obeys Strawson single-premise closure—because in a context where (18) and (17) both have truth values, the truth of (18) entails the falsity of (17). One such context is the context whose prospect set is just above—the prospect set that includes a *free-trip-on-the-Concorde* prospect. In this context, (18) is true and (17) is false.

Some have thought the case of (17) and (18) a counterexample to single-premise closure. But on my account, Strawson single-premise closure *explains* the case. (17) is true in one context and (18) is true in another, but when they are conjoined in (19)—and evaluated in a single context where they both have truth values—Strawson single-premise closure dictates that a contradiction results.

Notice that we have another instance of a difference in what an agent is truly said to want without any difference in his mental states. Across the contexts in our case, Andy is the same: he values a free flight on the Concorde, disvalues a paid one, and believes only a paid one is available. Andy is not truly said to want to take a flight on the Concorde in one context—e.g. one where (17) is true. In other contexts, Andy *is* truly said to want to take a flight on the Concorde. For example, we just saw a context—the one where (19) is evaluated—where (17), i.e. the negation of 'Andy doesn't want to take a flight on the Concorde', is false.

3.4.3 Single-premise closure: round three

In §3.4.1 I gave two examples, repeated below, to motivate single-premise closure:

(12) a. Andres wants to pursue a PhD in anthropology.
b. $\rightsquigarrow$ Andres wants to pursue a PhD.

(13) a. James wants to eat his dessert and relish every bite.
b. $\rightsquigarrow$ James wants to eat his dessert.

These inferences look unimpeachable. In fact, these inferences look so good that one wonders: if (12-a) is true in a context, how could (12-b) fail to be true in *any* context? (And likewise for (13-a) and (13-b).) One wonders whether Strawson-closure is strong enough to explain the inferences.

But in some cases where (12-a) is true in one context, (12-b) is indeed false in another. (And likewise for (13-a) and (13-b), although I won't show that here.) The differences between these contexts can be explained by a difference between prospects sets.

Imagine that Andres will pursue a graduate degree, but it's undecided whether he will go to law school, pursue a PhD in anthropology, or pursue a PhD in microbiology. Andres slightly prefers a PhD in anthropology to a law degree; both would be good, while pursuing a PhD in microbiology would be horrible. But it is not entirely up to Andres what graduate degree he will pursue; his father has a say in the matter. In particular, his father will allow Andres to choose between going to law school or pursuing a PhD, but if Andres elects for a PhD, his father will decide what discipline the PhD will be in. Andres is extremely confident that if his father decides Andres will pursue a PhD, his father will pick the PhD in microbiology. We may imagine the point at which Andres's father comes to Andres for the decision between law school and a PhD:

(20) a. [Father:] Do you want to go to law school or to pursue a PhD?
b. [Andres:] I want to go to law school, not pursue a PhD (it's too likely that if I went the PhD route, you'd pick a PhD in microbiology).

Despite the fact that (12-a) is true in one context in this case, (12-b) is false in another: namely, in the context in which (20-b) is true. Another way to bring out the falsity of (12-b) is to consider the more stilted (21):

(21) [Andres:] I don't want it to be the case that I pursue a PhD (since it's too likely that if that is the case, the PhD will be in microbiology).

My view explains how we can get a true reading of (12-a) in one context and a false reading of (12-b) in another. To see why, assume—as it's natural to do, given the setup of our case—that Andres assigns a high expected value to law school, a slightly higher expected value to a PhD in anthropology, and a very low expected value to a PhD in microbiology. It will then follow, given Andres's high confidence that his father would pick a PhD in microbiology over a PhD in anthropology, that Andres assign a low expected value to pursing a PhD, full stop. Consider the following two prospect sets:

Which-field? prospect set		
law degree	anthropology PhD	microbiology PhD

Which-degree? prospect set	
law degree	PhD

Assume for concreteness that we are interested in all-things-considered desire in evaluating (12-a) and (12-b) relative to these prospect sets. In a context with the *Which field?* prospect set, (12-a) is true: the prospect to which Andres assigns the highest expected value is the one in which he pursues a PhD in anthropology. But in a context with the *Which-degree?* prospect set, (12-b) is false: the prospect to which Andres assigns the highest expected value is the law degree prospect. It is natural to think that when we hear (20-b) as true in the conversation—i.e. when we hear (12-b) as false—the *Which-degree?* prospect set is exactly the one that is at issue. This is because Andres's father is effectively asking Andres to make a decision, and the decision Andres faces is represented by the *Which-degree?* prospect set: his decision is between between law school and a PhD. Which field a PhD would be in is not something about which Andres can decide.

We have yet another instance of a difference in the truth of a desire ascription without any difference in the agent's mental state. (12-b) is false, in the form of (20-b) in Andres's conversation with his father, but it is true, in the form of (22-b), in a conversation that Andres might have with a friend:

(22) a. [Friend:] I want to pursue a PhD. I'd love to study philosophy.
b. [Andres:] I want to pursue a PhD too. I'd love to study anthropology, but if I did decide to pursue a PhD, my father would probably have me study microbiology.

All of this is captured with my Desiderative Lockean view. As we've seen, the view explains the differences across contexts in this case solely by hypothesizing differences in the prospect set; nowhere did we posit any differences in the decision-theoretic terms with which we describe the agent's mental states.[15]

[15] With the data in §3, I've motivated a Desiderative Lockean view with considerable context-sensitivity. One might wonder whether these data can be explained by non-Desiderative Lockean views. I believe the answer is *yes*—so long as the views also have considerable context-sensitivity.

To see why, consider a particular non-Desiderative-Lockean framework, von Fintel's. His view, roughly, is: ⌜S wants p⌝ is true in c iff p is true in all of the worlds that S most prefers within her belief set. This view is a non-starter because it doesn't allow for context-sensitivity in 'want'. We might then generalize the view, in a way explored by Phillips-Brown (2018): S's preferences are evaluated not among the worlds in her belief set, but rather among the worlds in a contextually variable domain, Dom_c. The resulting view is what Phillips-Brown ms calls:

Generalized von Fintelian Semantics (GFS)
⌜S wants p⌝ is true in c iff p is true in all of S's most preferred worlds in Dom_c.

GFS cannot explain the data in §3.4.3, but it can explain the other data in §3. For reasons of space, I can only briefly discuss GFS's success with the data in §3.2 and §3.3, its failure with the data in, §3.4.3, and how it might be changed to accommodate the data in §3.4.3.

In §3.2, we encountered (8), 'Mo wants to win the race'. Intuitively, (8) is true in a context where we're considering Mo winning fairly, and false in a context where we're considering Mo winning by cheating. GFS can capture this by positing variation in the *Mo-wins* worlds that Dom_c contains. Roughly, if Dom_c does not contain *Mo-wins-fairly* worlds but does contain *Mo-wins-by-cheating* worlds, (8) is false; if Dom_c contains *Mo-wins-fairly* worlds, (8) is true.

So, GFS explains the data in §3.2 by allowing the p worlds in Dom_c to vary. GFS can likewise explain the data in §3.3 by allowing the $\neg p$ worlds in Dom_c to vary. But the data in §3.4.3 cannot be explained in either of these ways, which are the only ways for GFS to predict context-sensitivity in 'want'. Recall that (21), 'I don't want it to be the case that I pursue a PhD', is true in one context but false in another. My view explains this difference with different *partitions* of the space of relevant worlds. In the context where (21) is false, worlds in which Andres pursues a PhD in anthropology are "separated"—into different prospects—from worlds in which he pursues a PhD in microbiology, as we see in the *Which-field? prospect set*:

Which-field? prospect set		
law degree	anthropology PhD	microbiology PhD

But in the context where where (21) is true, these worlds are partitioned together into the proposition (prospect) that Andres pursuwes a PhD:

Which-degree? prospect set	
law degree	PhD

GFS cannot explain the data with different partitions; the view deals only in worlds, not how they are grouped together. We might then alter GFS so the domain is not a set of worlds but rather a partition on a set of worlds, PDom_c:

Partition-based Generalized von Fintelian Semantics
⌜S wants p⌝ is true in c iff p is *entailed* by all of S's most preferred *propositions* in PDom_c.

This view, which can predict the data in §3.4.3., is thoroughly context-sensitive: there must be context-sensitivity both in the p worlds and in the $\neg p$ worlds that the elements of PDom_c contain (to explain the data in §3.2 and §3.3), and in how these worlds are partitioned (to explain the data in §3.4.3).

References

Anand, P. and Hacquard, V. (2013). Epistemics and attitudes. *Semantics and Pragmatics*, 6(8):1–59.

Asher, N. (1987). A typology for attitude verbs and their anaphoric properties. *Linguistics and Philosophy*, 10(2):125–197.

Bermúdez, J. L. (2009). *Decision Theory and Rationality*. Oxford University Press.

Blumberg, K. (2023). Wishing, decision theory, and two-dimensional content. *Journal of Philosophy*, 120(2):61–93.

Blumberg, K. and Hawthorne, J. (2022). Wanting what's not best. *Philosophical Studies*, 179:1275–1296.

Cariani, F. (2013). 'Ought' and resolution semantics. *Noûs*, 47(3):534–558.

Crnič, L. (2011). *Getting* even. PhD thesis, Massachusetts Institute of Technology.

Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy*, 60(23):685–700.

Davis, W. (1984). The two senses of desire. *Philosophical Studies*, 45(2):181–195.

Drucker, D. (2017). Policy externalism. *Philosophy and Phenomenological Research*, 94(3).

Easwaran, K. (2016). Dr. Truthlove or: How I learned to stop worrying and love Bayesian probabilities. *Noûs*, 50(4):816–853.

von Fintel, K. (1999). Npi licensing, strawson entailment, and context dependency. *Journal of Semantics*, 16(2):97–148.

von Fintel, K. and Heim, I. (2011). Intensional semantics. Unpublished class notes.

Foley, R. (1992). *Working Without a Net: A Study of Egocentric Epistemology*. Oxford University Press.

Grano, T. and Phillips-Brown, M. (2022). (counter)factual want ascriptions and conditional belief. *Journal of Philosophy*, 119(12):641–672.

Grant, L. and Phillips-Brown, M. (2020). Getting what you want. *Philosophical Studies*, 177(7):1791–1810.

Heim, I. (1992). Presupposition projection and the semantics of attitude verbs. *Journal of Semantics*, 9(3):183–221.

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

Jerzak, E. (2019). Two ways to want? *Journal of Philosophy*, 116(2):65–98.

Karrttunen, L. (1973). Presuppositions of compound sentences. *Linguistic Inquiry*, pages 169–193.

Kratzer, A. (1991). Modality. In von Stechow, A. and Wunderlich, D., editors, *Semantics: An international handbook of contemporary research*, pages 639–650. de Gruyter.

Lassiter, D. (2011). *Measurement and Modality: the Scalar Basis of Modal Semantics*. PhD thesis, New York University.

Levinson, D. (2003). Probablistic model-theoretic semantics for *want*. *Semantics and Linguistics Theory (SALT) 13*, pages 222–239.

Navarro, P. E. and Rodríguez, J. L. (2014). *Deontic Logic and Legal Systems*. Cambridge University

Press.

Phillips-Brown, M. (2018). I want to, but.... In Truswell, R., Cummins, C., Heycock, C., Rabern, B., and Rohde, H., editors, *Sinn und Bedeutung 21*, pages 951–968.

Phillips-Brown, M. (2021). What does decision theory have to do with wanting? *Mind*, 130(518):413–437.

Phillips-Brown, M. (ms). Some-things-considered desire. https://semanticsarchive.net/Archive/mVhMTUyO/Some-things-considered%20desire.pdf. Unpublished manuscript.

Prior, A. N. (1958). Escapism: The logical basis of ethics. In Melden, A. I., editor, *Journal of Symbolic Logic*, pages 610–611. University of Washington Press.

van Rooij, R. (1999). Some analyses of pro-attitudes. In de Swart, H., editor, *Logic, Game Theory and Social Choice*, pages 534–548. Tilburg University Press.

Rubinstein, A. (2017). Straddling the line between attitude verbs and necessity modals. In *Modality Across Syntactic Categories*, pages 610–633. Oxford University Press.

Villalta, E. (2008). Mood and gradability: An investigation of the subjunctive mood in Spanish. *Linguistics and Philosophy*, 31(4):467–522.

Wrenn, C. (2010). A puzzle about desire. *Erkenntnis*, 73(2):185–209.