What it means to want

by

Milo Phillips-Brown

Submitted to the Department of Linguistics and Philosophy
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2019

Signature redacted

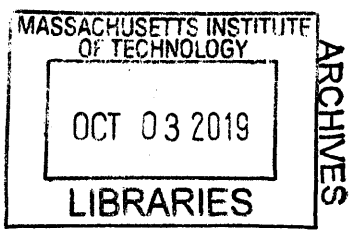
Author . . . . . .
Department of Linguistics and Philosophy
September 3, 2019

Signature redacted

Certified by . . . . . .
Stephen Yablo
David W. Skinner Professor of Philosophy
Thesis Supervisor

Signature redacted

Accepted by . . . . . .
Bradford Skow
Chair, Committee on Graduate Students



What it means to want

by

Milo Phillips-Brown

Submitted to the Department of Linguistics and Philosophy
on September 3, 2019, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Wanting is an easy concept to use. Talk to any three year-old and you'll know they've mastered it. Wanting is important, too. We understand one another in no small part through what they want, and wanting is a pillar in theories of mind and ethics. An account of wanting, then, must do dual duties: be powerful enough to carry this theoretical burden and simple enough to explain wanting's effortless use in daily life. The first two Chapters of this dissertation discharge these duties in part. The latter two Chapters complicate the task of discharging them further.

Chapter 1. Folk psychology and decision theory both represent our belief-like and desire- and preference-like states. Both use these representations to explain and predict our actions. If we can't account for one in terms of the other, we'd have a dubious dualism—two competing systems of representation, prediction, and explanation. I give a decision-theoretic account of a key folk-psychological notion—wanting.

Chapter 2. What we want depends on what we believe. Yet you can want to stay home (it would be nice to) despite believing it would ruin your career. This case confounds my theory from Chapter 1, as well as the orthodox semantics for 'want'. In Chapter 2, I develop a semantics based idea that you want to stay home *considering* its benefits, but *ignoring* the career consequences.

Chapter 3. The meaning of anankastic conditionals—like 'if you want to go to Harlem, you have to take the A train'—is clear, yet how it arises compositionally has proven an enigma. Many had thought the enigma unraveled by Condoravdi and Lauer (2016). I argue not: anankastic conditionals are still a mystery.

Chapter 4 (co-authored with Lyndal Grant). The widely held *Satisfaction-is-Truth Principle*—if A wants p, then A has a desire that is satisfied in exactly the worlds where p is true—posits an appealingly straightforward link between what we want and the satisfaction conditions of our desires, and in turn, enables appealingly straightforward accounts linking what we want, the wanting relation, and the contents of desires. We argue that the principle is nonetheless false.

Thesis Supervisor: Stephen Yablo
Title: David W. Skinner Professor of Philosophy

Acknowledgments

So many peopled helped me write this dissertation and grow as a philosopher during my time at MIT. Kyle Blumberg, David Boylan, Alex Byrne, Nilanjan Das, Tom Grano, Cosmo Grant, David Grant, Sally Haslanger, Jay Hodges, Magda Kaufmann, Justin Khoo, Matt Mandelkern, Kieran Setiya, and Steve Yablo: thank you. Thank you especially to Marion Boulicault, Kai von Fintel, Lyndal Grant, Abby Jaques, Ginger Schultheis, and Jack Spencer—you've changed how I think, and that's a wonderful thing.

Contents

Chapter 1

What does decision theory have to do with wanting?

1.1 Introduction

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. That is, both decision theory and folk psychology claim to show how our belief-like and desire- and preference-like states make sense of how we 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 much 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 folk-psychological notions such as believing and wanting. These notions have also been fruitful for studying both the mind (e.g. Davidson (1963); Dretske (1988)) and ethics (e.g. Smith (1994)), and some have even argued that folk psychology is essential to our very notion of personhood (e.g. Lewis (1974)).[1] Decision theory is similarly significant. It's played a key role in the development and practice of the social sciences, especially economics, since World War II (Erickson et al., 2013); decision theory is widely used in other disciplines, too, like neuroscience and philosophy.

In this paper, I will investigate whether we can account for folk psychology with decision theory. There are two cornerstone folk psychological notions—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

[1]The idea of a 'folk psychology' representing 'everyday' psychological notions raises pressing questions. Who are the folk of folk psychology? Whose daily lives does folk psychology structure? Whose concept of personhood is at stake? Issues for another another day.

confidence) and preference.

Many theorists have engaged in this task of accounting for belief. They've proposed necessary and sufficient conditions, stated in terms of credence, for when you're truly said to believe (e.g. Foley (1992), Hawthorne (2009), Easwaran (2016)).[2] (The *Lockean Thesis* says that there are such conditions.) The same accounting task for wanting has received much less attention. It is the task that I take up. I will give necessary and sufficient conditions, stated in terms of decision theory, for when you're truly said to want.[3] In other words, I provide an analogue of the Lockean Thesis for wanting.

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

Before we consider particular decision-theoretic accounts of wanting, we need to identify the concept or concepts of decision theory that we will use in formulating the accounts.

Decision theory's basic concepts are the credence function and the value function, which assigns a real number to each outcome. As sometimes understood, the value function represents an agent's non-instrumental preferences. The particular values it assigns are in and of themselves immaterial; what matters is the relationship between the values.[4] 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.

A tempting thought is that the value function can be the sole decision-theoretic concept in an account of wanting. It can't. As I just mentioned, the value function represents your *non-instrumental* preferences; and on the decision-theoretic understanding, your non-instrumental preferences are independent of what you believe. Of course, though, you want some things because you believe that they're instruments to other things you want, which is to say that some of your wants are *instrumental* (and some of your preferences are instrumental too). A purely non-instrumental concept, like the value function, can't furnish a full account of a partly instrumental concept, like wanting.

A case will bring this out. O'Neal will give his next paycheck to fight malaria. To keep things simple, imagine that you and I both non-instrumentally

[2]Others argue against these conditions: for example, Ganson (2008), Buchak (2014), Staffel (2016).

[3]There are other things to do, like addressing, as Dreier (2005) has done, Pettit's (1991; 2002) objection that decision theory doesn't have the right *desiderative structure* to account for wanting. Also important is determining whether our desire- and preference-like states are fundamentally one-place (like wanting) or two-place (like preference), a question taken up by e.g. Pollock (2006).

[4]Exactly which value function we pick is itself also immaterial. Any function that preserves the relative distances between the values will do.

value just one thing: the number of lives saved as a result of the donation. No other feature of an outcome makes a difference to our non-instrumental preferences. If in one outcome, O'Neal's donation results in more lives saved than in another, we each non-instrumentally prefer the first to the second. Further, the extent to which we each non-instrumentally prefer one outcome to another exactly matches the extent to which the number of lives saved in the first exceeds those in the second. In other words, our value functions match. But what we want O'Neal to do may not.

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. What we want is not the same, even though our value functions are the same. That's because although our non-instrumental preferences match, our beliefs, and so our instrumental wants and preferences, do not. I want O'Neal to donate to the Nets Foundation because I believe it's an instrument to saving lives; you don't have this belief and so you don't want him to donate.

The decision-theoretic concept in our account of wanting must, then, be sensitive to our belief-like states. Decision theory affords us many notions that are. The one I'll work with is *expected value*. It's a function of an agent's value function *and her credence function*.

The basic idea of expected value is this. Because I'm confident that the Nets Foundation is effective, I expect that a donation will likely result in a valued outcome, one where lives are 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 donation. 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. And this isn't merely coincidence. To repeat more or less from above: 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 isn't quite right for an account of wanting. That's because, independent of its relationship to wanting, expected value has been shown to be inadequate to represent non-ideally rational agents. However, a great deal has been done to correct the inadequacies.[5] And so the decision-theoretic concept for a proper account of wanting will be an entity like expected value that reflects these corrections.[6] To keep things simple, though, I'll work with expected value in what follows.

[5]Buchak (2013) and Chandler (2017) review both the inadequacies and the corrections.

[6]Worth noting are cases where an agent wants something that she is positive will not happen (for example, I want to live forever, but I'm sure that I won't), cases that preclude using certain definitions of expected value to account for wanting. Wrenn (2010) solves the problem.

1.3 Against What's-best Accounts

We've identified the 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 (1996, 2005), Villalta (2008), Rubinstein (2017)).[7] Most What's-best Accounts that have been offered aren't decision-theoretic—few decision-theoretic accounts have been proposed[8]—but we can translate them into our decision-theoretic framework.

Consider a few What's-best Accounts.[9]

> S wants p iff S prefers p to not-p. (Davis, 1984)
>
> wanting something is preferring it to certain relevant alternatives, the relevant alternatives being those possibilities that the agent believes will be realized if he does not get what he wants. (Stalnaker, 1984, p. 89)
>
> ⌜S wants p⌝ is true iff p is true in all of the best worlds compatible with S's beliefs, as ranked by S's preferences. (paraphrasing von Fintel (1999))
>
> ⌜S wants p⌝ is true iff the expected value S assigns to p exceeds the expected value S assigns to not-p. (Levinson, 2003)

(Most of these accounts presuppose that wanting is a propositional attitude, a pedigreed, if contested (e.g. Montague (2007); Moltmann (2013)), presupposition. To keep things simple I will adopt 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 iff S assigns a higher expected value to p than to any of certain alternatives.

The Simple What's-best Account is false. Its problem is shared by more sophisticated What's-best Accounts, decision-theoretic and otherwise: being best is

[7] What's-best Accounts are also the standard among those who understand *conditional* wanting in terms of preference (e.g. Edgington (1995)).

[8] van Rooij (1999), Levinson (2003), Wrenn (2010), and Jerzak (2019) offer decision-theoretic What's-best Accounts.

[9] '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.

neither necessary nor sufficient for being wanted. This becomes apparent when we consider certain mundane facts about wanting.

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.[10] (Maybe you can be truly said to want to shoot, on which more in section 1.7. 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. Imagine that you're going out to dinner. The 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 say, 'I want to go to the ramen shop, and I 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 Crnič (2011)—have argued that pizza–ramen-type cases are not in fact counterexamples to the thesis that being best is necessary for being wanted. They hypothesize that 'want' is *context-sensitive*. An agent is represented with many value functions, not just one, and which value function is used to assess 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 bread-y and cheesy, assigning higher values to outcomes to the extent that you eat things that are bread-y and cheesy in

[10] 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, 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.

them. Relative to this function, pizza is best, while ramen is best according to a function that favors eating broth-y soups with noodles.

The view is as follows. Where c is a context:

> *What's-best Account with Varying Value Functions*
> ⌜S wants p⌝ is true in c iff S 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 these sentences to be true:

> *Pizza.* I want to go to the pizzeria.
>
> *Ramen.* I 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 bread-y and cheesy, while *Ramen* comes out true against a value function that represents your interest in things noodle-y and broth-y. 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 blue ticket (worth 100 dollars) and the red ticket (worth 50 dollars). You might say: 'I want to get the red ticket. And I *really* 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 ticket. 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.[11]

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

1.4 A simple What's-good-enough Account

What's-best Accounts are wrong. 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 count as what I'm calling 'What's-good-enough Accounts'

[11]Thank you to Kieran Setiya and Rob Pasternak for help developing this case.

have been proposed—none are motivated in the way that I motivate mine. I address them in footnotes 12 and 20.)

What's-good-enough Accounts claim that you want, well, what's good enough (in your eyes). This corrects the mistakes 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 iff the expected value S assigns to p meets a certain threshold.[12]

Compare to a simple version of the Lockean Thesis: ⌜S believes p⌝ is true iff the credence S 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 virtue: 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?

On What's-good-enough Accounts, cases of directly conflicting desires couldn't be more simple: 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

[12]van Rooij floats a Simple What's-good-enough Account: roughly, ⌜S wants p⌝ is true iff the expected value S assigns to p exceeds the expected value that S assigns to a tautology (he doesn't explain why he uses a tautology for the threshold). Pollock (2006) proposes that you want something just in case it's both best *and* good enough (although he doesn't put it in those terms).

be good enough; both can be wanted.

The facility of What's-good-enough Accounts with 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 are no problem.[13]

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.

1.5 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 is 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 onlooker hearing you ask me to drive to the pizzeria. He could say, 'I guess she [i.e. you] ultimately does not want to go to the ramen shop'. In this context, *Ramen* is false.[14])

[13]McDaniel and Bradley (2008) also argue against preference-based accounts of wanting—in particular, of *conditional* wanting—while presupposing a kind of What's-best Account. ('The preferentist analysis of conditional desire is: S desires P given Q iff S prefers P & Q to $\neg P$ & Q' (p. 283). This is Edgington's (1995) view.) They claim that these accounts are extensionally inadequate, and they may well be right. But even if they are, it wouldn't follow that What's-good-enough Accounts of conditional wanting are extensionally inadequate. (I don't develop such an account here.)

[14]More precisely, *Ramen* is false if the (clunky) *non*-neg raising of the onlooker's sentence

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;[15] it's context-sensitive.[16]

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 in case some condition C obtains, where C is insensitive to context. To take just some examples (philosophical encyclopedia entries on desire are littered with others):

> Repeated from above: *Simple What's-best Account*
> ⌜S wants p⌝ is true iff S assigns a higher expected value to p than to any of certain alternatives.
>
> Repeated from above: *Simple What's-good-enough Account*
> ⌜S wants p⌝ is true iff the expected value S 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.[17]
>
> S wants p iff S is disposed, other things equal, to do what she believes will bring it about that p obtains.[18]
>
> S has an 'intrinsic appetitive desire that P' iff S 'constitute[s] P as a reward'. (Arpaly and Schroeder, 2013, p. 128)

Read in a straightforward way, 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).

---

is true: 'I guess she [i.e. you] ultimately isn't such that she wants to go to the ramen shop'.

[15]This conclusion won't come as a surprise those who think that *mental content* isn't intrinsic to you, and so that what you're truly said to want is thereby not intrinsic to you (see e.g. (Brown, 2004)). The kind of non-intrinsic-ness I'm talking about here—due to context-sensitivity—is different. It cross-cuts the debate about the intrinsic-ness of content.

[16]Note that various authors have argued—appealing to different data than mine—that want ascriptions are context-sensitive: Levinson (2003), Villalta (2008), Crnič (2011), Lassiter (2011), Condoravdi and Lauer (2016), Phillips-Brown (2018), Dandelet (ms); Jerzak (2019) says 'want' is assessment-relative.

[17]Morillo (1990), among others, proposes a view like this.

[18]Stalnaker (1984, p. 15), for example, advocates a principle in this vein.

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 right. 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:[19] 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 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* could have shifting truth values. It 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 1.7, 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 really *want* to go to. (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.

1.6 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 propose that what you're truly said to want is what's good enough—I propose a What's-good-enough Account—and that what shifts by context is *what counts as good enough.* Or, formally, what shifts is the threshold.

> *What's-good-enough Account with Varying Threshold*
> ⌜S wants p⌝ is true in c iff the expected value S assigns to p meets *the threshold in* c.[20]

[19] You might think, on other grounds, that what you're disposed to do is not intrinsic to you.

[20] Two others give accounts that can be characterized as What's-good-enough Accounts with Varying Threshold. Bradley (1999), whose concern is conditional desire, says in passing

Compare to a context-sensitive version of the Lockean Thesis: ⌜S believes p⌝ is true in c iff the credence S assigns to p meets the threshold in c (e.g. Hawthorne (2009)). Further, note that on my view, '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.[21]

My account shares the merits of the Simple What's-good-enough Account, and as we'll now see, corrects its failure in cases of context-sensitivity.

Recall that we had three contexts: one where *Pizza* and *Ramen* are both true, one where just *Pizza* is true, and one where neither are. Formally, my account says 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 is higher still where both sentences are false. Neither the ramen shop nor the pizzeria meet it. All of our cases are covered.

Let's now say more about 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.

Start with the context where you say, repeating now from above, 'of course neither of these places do I really *want* to go to. (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 restaurant that, to your mind, is of a certain quality—the kind of restaurant that you could find in a bigger city than the one that you live in. 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

---

that '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). Bradley doesn't motivate his view or elaborate on it. Lassiter (2011) says that ⌜S wants p⌝ is evaluated with respect to a set of alternatives. This set, which varies by context, determines a threshold, and as the set varies with context, so too does the threshold. His view is that ⌜S wants p⌝ is true in c iff the expected value S assigns to p is 'significantly greater' (p. 182) than the *average* of the expected values that S assigns to the relevant alternatives in c. (The threshold, then, is some particular value that's significantly greater than this average.) For Lassiter, there is no shift in truth value without a shift in alternatives. As we've seen, though, we get truth-value shifts even holding the alternatives fixed—ramen shop, pizzeria, hot dog stand.

[21] See e.g. Hawthorne (2009) on 'believe' within the Lockean literature, and Stalnaker (2008) without; e.g. Stine (1976) on 'know'; Villalta (2008) on the analogues of 'fear', 'hope', and 'glad' in Spanish; Blumberg and Holguín (ms) on 'surprise' and 'suspect'; Pearson (ms) reviews further literature on this topic.

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. In other words, both shooting and not look horrible in comparison to the kinds of actions you'd choose to have available.[22]

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 me to instruct you where to go; it would stand to reason that you should be instructed to go only where I most prefer to go, which is why what counts as good enough is only what I most prefer.[23]

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 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 considering whether, and if so where, to go to dinner together. I hypothesize that what counts as good enough is anything that's better, in your eyes, than staying 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.

[22]Daveney (1961) also discusses, at length, how we use 'want' to communicate how we would choose to act when under duress versus not.

[23]Things are more complicated when there's a tie in what you most prefer. I'll leave this issue for another day.

1.7 On multiple-senses hypotheses

I've traced the context-sensitivity of want ascriptions to *context-sensitivity in 'want'*. As I mentioned in section 1.5, others trace 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. Some have denied the phenomena, and my account is compatible with either their denial or their acceptance. I'll illustrate all of this with two commonly discussed senses of 'want': *all-things-considered* 'want' and *volitive* 'want'.

(Note that we should be skeptical of multiple-senses hypotheses. (This does *not* require skepticism of the phenomena that they purport to explain.) 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 is ambiguous.[24] And, as I showed in section 1.5, 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 ones on which it at least three-way ambiguous, an especially undesirable result.)

What you want, all things considered, is normally taken to be what you prefer most, given certain alternatives. 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 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

[24]Except in cases where other languages lexicalize the term differently—e.g. how 'bank' can translate into German as either 'Bank' or 'Ufer'. To my knowledge, this is not the case with 'want'.

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 their disagreement—in terms of whether the threshold has a floor.

Turn now to the controversial *volitive* sense of 'want'.[25] 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 context-sensitivity in 'want'.[26]

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.[27] 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 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.

1.8 Conclusion

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

[25]Davis (1984) reviews the literature.

[26]Levinson (2003) and Condoravdi and Lauer (2016) propose to do the same, but with different accounts than mine. They both endorse the thesis that you can always be truly said to want what you intend.

[27]However, without getting into details, my view is not compatible with a certain stronger version of the idea plus a commitment to a special kind of Buridan's Ass case.

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 proposed 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.

There are still problems to be addressed, problems that have been raised by others, and which I leave to a footnote. None are directly targeted to accounts like mine (decision-theoretic What's-good-enough Accounts), but each can be modified to be. Each has a proposed solution, and each solution, I hope, can also be modified to assist accounts like mine, but that remains to be seen.[28]

Even without definitive solutions to these problems, we have a lot to work with. What's-best Accounts, the orthodox accounts that link wanting with preference, we now know are wrong. Their mistakes are corrected by What's-good-enough Accounts. What's-good-enough Accounts also cleanly handle the pervasive but theoretically puzzling phenomenon of directly conflicting desires, which had been thought to falsify preference-based accounts of wanting.

In line with the more general project on which attitude verbs are context-sensitive, I have proposed that we add a contextually variable bar for what counts as good enough. This empowers us to understand 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, make sense of the phenomena that have motivated traditional multiple-senses hypotheses without committing to an unsavory ambiguity.

[28]First, Jackson (1985) and Pettit (1991, 2002) hypothesize that wanting can take either prospects or properties as objects. Pettit charges that decision theory countenances only prospects; Dreier (2005) replies. Second, there is Villalta's (2008) *doxastic problem* (named by Rubinstein (2017)); Crnič (2011) outlines one solution; Rubinstein (2017) offers another; in Chapter 2, I develop my own solution. Finally, Jerzak (2019) introduces problem data of what he calls the *advisory* 'want', and also develops his own solution.

Chapter 2

Some-things-considered desire

2.1 Introduction

It's a plain fact of life: what we want depends on what we believe. You want to stay home from work today, and you want this because you believe that an agreeable prospect will come about: you'll be able to relax and work on that painting you never seem to have time for. Yet you also believe that if you stay home, something you greatly disvalue will follow: you'll miss the biggest meeting of your life, and your career will be ruined. (Because of this, you intend not to stay home.) This is a unexceptional description of a familiar phenomenon. Yet it sparks a question: you want to stay home because of what you believe will follow, yet you greatly disvalue what you believe will happen, all told, if you stay home—how can this be?

This question poses not just a puzzle but a crisis for the dominant approach to the semantics of 'want', which attempts to capture the connection between wanting and believing. The approach, motivated not just by this obvious connection but also by powerful theoretical considerations, is adopted by the standard-bearers of the literature on 'want'—Heim (1992), von Fintel (1999), and Levinson (2003)—among many others, including me in Chapter 1. The idea, viewed from 30,000 feet, is that you're truly said to want p just in case you have a certain positive attitude towards what you believe will happen if p obtains.[1] (Authors differ on what this positive attitude is, among other things.) Yet you want to stay home despite having a negative attitude towards what you believe will happen if you do.

An immediate response might be that there's really no crisis at all: it isn't in fact true that you want to stay home today, in light of what you believe the consequences to be. This response is half right. There are indeed contexts where it's not true to say that you want to stay home. (You might be asked if you want a ride to work or if you want to stay home, and reply 'No, I'll take the ride, since my career will be ruined if I stay home.') Nonetheless, there are also contexts where it's true to say that you want to stay home. (You might

[1] It's taken for granted that wanting is a propositional attitude, but whether that's so is orthogonal to the problem I raise.

say, 'I want to stay home, but of course I don't intend to because it will end in disaster.') This is all we need to generate the problem: a context where it's true to say that you want to stay home. And so the question remains. How can it truly be said that you want something that you believe will result in an outcome that, all told, you greatly disvalue?

To answer this question while preserving the connection between wanting and believing in our semantics, we must re-imagine the cases at hand. They involve what I call *some-things-considered, other-things-ignored* desire. Considering what you believe to be the benefits of staying home, but ignoring what you believe about the consequences for your career, you're truly said to want to stay home. Considering both the benefits of staying home and the consequences for your career, you're not truly said to want to stay home.

2.2 The problem

As I said in the Introduction, the dominant approach to says, in outline, that you're truly said to want p just in case you have a certain positive attitude towards what you believe will happen if p obtains—this is how wanting and believing are supposed to be connected. This approach has been shown to make correct predictions in a wide range of cases. It also helps explain the relationships between 'want', 'wish', and 'be glad';[2] various issues raised by Crnič (2011, Appendix A); the distinction between what Jerzak (2019) calls the *predictive* 'want' and the *advisory* 'want'; puzzles raised by Blumberg (2018) and Blumberg (ms);[3] and how 'want' interacts with conditionals.[4]

The dominant approach has been adopted in different ways. For example, theorists disagree about what the relevant positive attitude is. Some say that it's a matter of preferring what you believe will happen if p obtains to what you believe will happen if not-p obtains.[5] (Such views I call 'What's-best Accounts' in Chapter 1.) Others maintain that you're truly said to want p just if what you believe will happen if p obtains is 'good enough' in your eyes (this being the positive attitude)—just if it meets some threshold or baseline.[6] (Such views I call 'What's-good-enough Accounts' in Chapter 1.) The various semantics differ in other ways, too—e.g. whether the relevant notion of belief is all-out or graded.[7]

[2]See Heim (1992) and Grano and Phillips-Brown (ms).

[3]Blumberg's puzzles concern 'wish' (his 2018) and 'hope' (his ms), but they can be generated with 'want' as well.

[4]See Jerzak (2019) and von Fintel (ms).

[5]For example, Heim (1992), von Fintel (1999), Levinson (2003), and Jerzak (2019) on the predictive use of 'want'. Note that the problem also arises for what Jerzak calls the advisory use of 'want': according to Jerzak, the truth value for a desire ascription with one use (either predictive or advisory) can come apart from the other use when the information state of the agent differs from the information state in the context of assessment. In our cases, though, we can simply assume that these information states are the same.

[6]For example, Lassiter (2011) and me in Chapter 1.

[7]For example, Heim (1992) and von Fintel and Iatridou (2005) go for all-out belief, while

The differences in how the dominant approach has been adopted don't matter in the face of the fact that you can truly be said to want p despite having a *negative* attitude towards what you believe will happen if p obtains. You can truly be said to want to stay home despite believing that if you do, your career will be ruined. You greatly disprefer what you believe will happen if you stay home to what you believe will happen if you don't (relaxing, painting, and ruining your career is much worse than neither relaxing nor painting but having an intact career). What you believe will happen if you stay home is nowhere near good enough in your eyes; it's awful.

I'll illustrate the problem with a case study: a simplified version of the best worlds semantics of von Fintel (1999), who adapts Kratzer's (1981, 1991) classic system of modals. As I mentioned above, the semantics belongs to the class of semantics that say that you're truly said to want p just if you prefer what you believe will happen if p obtains to what you believe will happen if not-p obtains. In particular, the truth conditions are these:[8]

> *Truth conditions*
> If defined, $[\![\text{A wants p}]\!]^c = 1$ iff, among the worlds that are compatible with A's beliefs, p is true in the best of them, as ranked by A's preferences.

Let's now state the semantics precisely. It has two contextually determined functions: a modal base f_c and ordering source g_c (c being a context). Both take an agent (and a time and a world, strictly speaking) as input; the modal base returns a set of worlds, while the ordering source returns a set of propositions. (Sometimes I'll use 'modal base' and 'ordering source' to refer to the functions and sometimes to their outputs.) The important thing to know about the ordering source is that it determines a ranking on worlds $\preceq_{g_c(A)}$, which ranks worlds on the basis of preferences of the agent A.[9]

The semantics put in words is this: if defined, $[\![\text{A wants p}]\!]^c = 1$ just if p is true in all of the best worlds in the modal base in c, as ranked by the ordering source in c (i.e. A's preferences). More formally, where best_{A,f_c,g_c} is the set of best worlds in $f_c(A)$, as ranked by $\preceq_{g_c(A)}$:[10]

> *Best-worlds Semantics*
> If defined, $[\![\text{A wants p}]\!]^c = 1$ iff $\forall w \in \text{best}_{A,f_c,g_c}\colon p(w) = 1$.

The modal base is where the agent's beliefs enter the picture: it is the set of worlds compatible with what the agent believes, her belief set.

> *Belief Set Modal Base*
> $[\![\text{A wants p}]\!]^c$ is defined only if the modal base in c, $f_c(A)$, is the set of

e.g. Levinson (2003) and I in Chapter 1 advocate for a graded notion.

[8] I adopt the following convention for variables: 'A' ranges over names, 'A' ranges over the corresponding agents denoted by those names, 'p' ranges over terms that denote propositions, and 'p' ranges over the corresponding propositions denoted by those terms.

[9] $w' \preceq_{g_c(A)} w''$ iff $\{p \in g_c(A) : p(w'') = 1\} \subseteq \{p \in g_c(A) : p(w') = 1\}$.

[10] $\text{best}_{A,f_c,g_c} = \{w' \in f_c(A) : \neg\exists w'' \in f_c(A), w'' \prec_{g_c(A)} w'\}$.

worlds compatible with A's beliefs.

Identifying the modal base with the belief set is a simplification. There are two kinds of cases where it clearly will not do. First, when the agent either believes p, in which case $[\![\text{A wants p}]\!]^c = 1$ vacuously, or she believes not-p, in which case $[\![\text{A wants p}]\!]^c = 0$ vacuously (Heim, 1992). And second, when we're in a context where the 'want' at issue is the advisory 'want', in which case it's not the agent's beliefs that matter, but rather the information in the context of assessment or certain other facts (Jerzak, 2019). We can safely set both kinds of cases aside.

Combining the Belief Set Modal Base with the Best-worlds Semantics yields the truth conditions from above. Those truth conditions are falsified by our case where you can truly be said to want to stay home, despite believing that doing so would result in disaster. The following sentence is true in some contexts:

Home. You want to stay home.

Home is predicted false. Because you believe that your career will be ruined if you stay home, all of the worlds in your belief set where you stay home are ones where your career is ruined. These worlds aren't best worlds—so, *Home* comes out false—and that's an understatement. These worlds are *awful.* In them your career is ruined!

The same problem arises in any case of this form, any case where you want p despite believing that, all told, p will lead to something that's awful in your eyes. Worlds in your belief set where p is true are therefore awful; you're predicted not to want p.[11]

The semantics is supposed to capture the relationship between believing and wanting via the modal base, but when the modal base is the belief set, the agent's beliefs exert an outsized influence. The truth of *Home* supposedly hinges on your attitude towards worlds in your belief set where you stay home. Yet in every one of these worlds, your career is ruined. Within your belief set, staying home is *entangled* with being financially ruined. When staying home is entangled with being financially ruined, staying home comes out looking horrible, since the good of staying home is greatly outweighed by the bad of career ruin.

2.3 What if different ordering sources represent different values of the agent?

What might be happening when you want what you believe will result in something awful? Maybe it's that the result, seen in a certain light, isn't awful after all. Sometimes we want various things because we have various values. You value career success, yes, but you also value relaxing and painting. When

[11]The exception will be in cases where you believe that things will be even worse if not-p obtains, in which case you'll be predicted to want p.

viewed in the light of this latter value, staying home, despite the consequences, isn't awful. It's good!

We can represent these notions of values of the agent within the semantics via the ordering source, which represents the agent's preferences: the ordering source in one context represents different of your values than the ordering source in another. This is what Levinson (2003) proposes (within a different semantics).[12] For example, the ordering source that represents your value of relaxing and painting favors worlds where you relax and paint over those where you don't. In your belief set, worlds where you relax and paint are those where you stay home. Such stay-home worlds are *best* in your belief set, relative to this ordering source. In a context that selects this ordering source, *Home* is true. Success—in this case, at least.

Our problem is that staying home is entangled with having your career ruined (because within your belief set, all of the staying home worlds are career-ruin worlds). With different values we can disentangle the two, in a sense. The very same worlds, the stay-home-career-ruin worlds, can be best relative to one value (one that favors staying home) and not best relative to another (one that favors an intact career).

It very well may be that different ordering sources represent different values of the agent. (Levinson uses the view that they do to solve a different problem, and in Section 9, I note that it may solve another.) Different values won't furnish a general solution to our problem, though. Imagine Jim, a bookie. Jim makes it a point to form no emotional attachments to his clients, and when it comes to interacting with them, he cares about one thing: how much money he himself has. Jim gets no special pleasure from getting money from any particular person or special displeasure of having to pay any particular person. Jim believes that he owes Miranda $100,000 and that Anna owes him $50. Miranda and Anna are always with one another: Jim believes he'll be paid by Anna if and only if he pays Miranda. Jim can truly be said to want to be paid by Anna. (He might say, 'I want to be paid by Anna, but of course I don't plan to get her to pay me back, because that would mean paying Miranda.') Yet worlds in his belief set where he's paid by her are awful, for in them he also repays Miranda, and loses $999,950!

This is one of our problem cases: the agent wants something despite believing that it will lead to a truly awful consequence. The multiple-values view hypothesizes that, viewed in the light of a certain value of Jim's, the consequence (losing $999,950) isn't awful after all. There's some ordering source that represents Jim's values that ranks worlds where he loses $999,950 above ones where he loses no money at all. Yet what would this value be? There is no such value of Jim's. In any interaction with his clients, it's money, and money alone, that matters to Jim. For Jim, losing $999,950 is awful no matter how you look at it.

[12]Crnič (2011) also proposes that the ordering source varies by context, although he doesn't say what the ordering sources are supposed to represent.

You might reply that that there must be something about the worlds in his belief set where Jim loses $999,950 that privileges them over the worlds in his belief set where he loses nothing at all. One thing that distinguishes these two kinds of worlds is that in the ones where he loses all this money, he's paid by Anna, and the ones in which he loses nothing, he's not paid by Anna. Could it be that relative to the value of being paid by Anna, worlds where she pays Jim are better than worlds where she doesn't pay Jim? Could it be that being paid by Anna in a world, despite the disastrous financial consequences, places it above a world where he's not paid by Anna? Not for Jim. Jim doesn't care about who he pays or is paid by. In client interactions, Jim's sole value is money. No ordering source that represents *his* values rates financially worse worlds over financially better ones (in cases where he's interacting with clients).

The lesson from this Section is that the agent's values can't always disentangle two things within the belief set, for sometimes there is only one relevant value at play. The agent's beliefs still loom too large.

2.4 What if the modal base isn't the belief set?

A simple way to reduce the role of belief is to admit worlds outside of the belief set into the modal base:

> *Beyond Belief Set Modal Base*
> $[\![\text{A wants p}]\!]^c$ can be defined even if $f_c(A)$ contains worlds incompatible with A's beliefs.

As when we introduced the Belief Set Modal Base, a caveat is in order. When the agent either believes p or believes not-p, or we are in a context where the advisory 'want' is at issue, the modal base cannot be the belief set. In these cases, we must adopt the Beyond Belief Set Modal Base. As before, we can set these cases to the side. In assessing the Belief Set Modal Base in this Section, we are considering cases not of these kinds.

If the modal base can contain a world where you stay home without having your career ruined, we can disentangle having your career ruined and staying home. And such a world may be best; *Home* may be true.

We should be skeptical of the Beyond Belief Set Modal Base. For one, going beyond the belief set may jeopardize some of the theoretical benefits that we gain from identifying the modal base with the belief set—whether it in fact does has not been explored. Further, we may wonder whether the Beyond Belief Set Modal Base compromises the maxim that what we want depends on what we believe, since it places the truth value of a 'want' ascription partly on the agent's attitude towards worlds that are incompatible with her beliefs. This worry is on display in problem cases that arise for those—Rubinstein (2017) and Anand and Hacquard (2013)—who have accepted the Beyond Belief Set Modal Base, or rather its analogue in other semantics.[13]

[13]Neither Rubinstein's nor Anand and Hacquard's views are Best-worlds Semantics.

Rubinstein suggests that the analogue of the modal base within her semantics may be a superset of the belief set: in cases similar to our own (see Section 8), it can contain worlds that are similar to the world w that the agent inhabits (or as Rubinstein puts it, worlds circumstantially accessible from w). These worlds may be completely divorced from the agent's beliefs, since the agent may be entirely wrong about what her world is like.[14]

Rubinstein's approach to the modal base is wrong, as Grano and Phillips-Brown (ms) point out. We can see the problem in our staying home case, where we would admit into the modal base worlds similar to the world w that you inhabit. Imagine that in w, unbeknownst to you, a secret assassin lies in wait, ready to poison all of your friends if you stay home. The same will hold in the worlds similar to w. Our goal is to find a modal base relative to which *Home* is true—a modal base in which the best worlds are stay-home worlds—and this modal base does not fit the bill. Stay-home worlds where your friends are poisoned are even worse than the stay-home worlds in your belief set, and those worlds were already not best!

We have strayed too far from the agent's beliefs. Worlds with secret assassins have no place in our semantics for *Home*, since they are irrelevant to you wanting to stay home. They're irrelevant because you believe, about as strongly as you believe anything, that there aren't assassins lying in wait. (Imagine saying that you want to stay home, and someone replying, 'But what if there are secret assassins?' A non-sequitur if there ever was one.)

According to Anand and Hacquard (2013), who follow a suggestion of Villalta's (2008), the analogue of the modal base in their semantics may greatly exceed the belief set. The semantics is contrastive: when we evaluate $[\![\text{A wants p}]\!]^c$, there is a contrast class, a set C of propositions, determined by c. The analogue of the modal base contains the p worlds and the union of the members of C. (So if C contains q and r, for example, then the modal base is the p worlds, the q worlds, and the r worlds.) Anand and Hacquard consider only cases where the contrast case contains simply not-p (p. 8), and I will follow their lead.

When the contrast class contains simply not-p, the modal base contains the p worlds and the not-p worlds—i.e. the modal base is the set of all worlds. If *Home* is to be true, certain worlds where you stay home must be better than every possible world where you don't stay home. This just can't be, for the possible ways for you not to stay home are too varied. There are worlds, for example, where all people of Earth are united under one nation and you are its

Yet if $[\![\text{A wants p}]\!]^c = 1$ on either of Rubinstein's or Anand and Hacquard's views, then $[\![\text{A wants p}]\!]^c = 1$ on the Best-worlds Semantics. So cases like the ones I present later in this Section—where a sentence that's true is predicted false by the Best-worlds Semantics—are counterexamples to Rubinstein's and Anand and Hacquard's views.

[14]Rubinstein also says that we should expand the modal base by 'suspending' (p. 118) some of the agent's beliefs, which is on the right track: the agent's beliefs need to constrain the semantics some, but not too much. What I argue against later in this Section is her way of suspending (by way of circumstantial accessibility).

monarch; when you walk the streets the adoring masses applaud you. That's at least as good as staying home. Needless to say, such outrageous worlds—such worlds so far from what you believe to be possible—are irrelevant to your desire to stay home. Again, we've strayed too far from the agent's beliefs.[15]

The through-line between these views that extend beyond the belief set is that they go wrong because they admit worlds into the modal base that are irrelevant to the agent's desires. (Recall that we're setting aside both those cases where the agent believes p or believes not-p and those cases where the advisory 'want' is at issue.) And this shouldn't be a great surprise, since what we want depends on what we believe.

2.5 Some-things-considered desire

Strong theoretical considerations recommend a semantics for 'want' on which the agent's beliefs play a leading role, but as we've seen, the semantics that have so far implemented this approach—in our case study, one that identifies the modal base with the belief set—makes that role too great.

To get an appropriately sized role, I propose a situation semantics based on what I call *some-things-considered, other-things-ignored* desire. When *Home* is true, when you're truly said to want to stay home, you are—in a sense—considering certain things but ignoring others. Considering everything that would be nice about staying home, but ignoring what the consequences for your career would be, what's good about staying home comes to the fore and what's bad about it recedes. Considering what for you would be good about staying home, but ignoring what would be bad, you want to stay home.

This paper started with a question: how can it be that we want what we believe leads to other things we value, yet you want to do something (stay home) that we believe leads to things that, all told, you greatly disvalue? The answer is that when we want something, all is not always told. Some things are considered and others are ignored. Although you believe that staying home will lead to a ruined career, this belief doesn't matter when the career effects of staying home are ignored. When this belief is sidelined—in a sense I make precise below—you want to stay home. Belief plays a role, but a limited one. You believe that you'll do things you enjoy if you stay home (that's why you want to stay home); you also believe that staying home would have awful consequences. The former beliefs, but not the latter, come into play when considering what it would be like to stay home but ignoring consequences.

The terms 'considering' and 'ignoring' might mislead in that they suggest that whenever an agent wants, some things considered and other things ignored, she is actively considering and ignoring (or has actively considered and ignored). It some cases we do actively consider and ignore—you could

[15]This argument applies to Anand and Hacquard's particular way of implementing the contrastive view (where not-p is the contrast proposition). We might implement it in another way, of course. I don't have a general argument against any possible implementation, but I can't see why we should expect that a contrastive semantics would help with our problem.

say, 'Considering what it would be like to stay home, but ignoring the consequences, I want to stay home'—but that needn't always be so. Often we want, some things considered and other things ignored, without having sat down and considered or ignored. Here, my use of 'considering' and 'ignoring' is metaphorical.

By using 'considering' and 'ignoring' in these ways—sometimes literally, sometimes metaphorically—I follow the use of another philosophical term about desire: the 'considered' of 'all-things-considered desire'. Sometimes we want, all things considered (whatever exactly that amounts to), after actively weighing all the pros and cons. Sometimes we want, all things considered, after considering all things. In such cases, 'considered' is literal.

But it's not always literal. It's often said that rational agents tend to do what they want, all things considered. And rational agents don't always sit down and consider all things before acting. Imagine that in driving home today, you turned on your left blinker before turning left onto your street. This you wanted to do, all things considered. However, you needn't have, and in normal circumstances wouldn't have, mapped out in the moment all the things that spoke in favor or against turning on your blinker. Here, 'considered' is metaphorical.

We can make precise the notion of some-things-considered desire within a formal framework, which will serve as the basis for our semantics. We'll work in three steps: what it is to consider some things but not others, how an agent's beliefs fit with some things being considered and other things ignored, and how desire enters the picture.

I model the notions of considering and ignoring with situations. A situation stands in contrast to a possible world in the sense that a possible world is maximal, while a situation is partial. We can think of a possible world as, or as determining, a classical model. I will think of a situation analogously: either as, or as determining, a *partial* classical model.[16] Some propositions are neither true nor false relative to a given situation, but rather undefined. (In the limit case, a situation relative to which every proposition is either true or false is simply a world.)

It is the partiality of situations that lends itself to representing ignoring. When a proposition is undefined relative to a situation, it is, in a sense, not represented: it is ignored. When a proposition is true or false relative to a situation, it is, in a sense, represented: it is considered. More rigorously, a proposition p is considered relative to a situation s just if p is either true or false at s; a proposition p' is ignored relative to s just if p' is undefined at s. By extension, p is considered (ignored) relative to a set of situations S just if p is considered (ignored) relative to each member of S.

An agent's beliefs enter the picture as follows. Before, we had the set of all worlds, and from that we derived the agent's belief set: the set of worlds compatible with her beliefs. Now, we have sets of situations relative to which

[16]For a similar approach, see Humberstone (1981) on what he calls 'possibilities'.

some things are considered and other things are ignored. Given any such set S, there is an analogue of the belief set: the set of situations in S compatible with what the agent believes.[17] Such a set represents what the agent believes, considering some things and ignoring others.

The crux of the problem we've been wrestling with is that within your belief set, staying home is entangled with a ruined career, which is to say that in your belief set, the worlds where you stay home are all ones where your career is ruined. We're empowered to disentangle the two with the notions of considering and ignoring. Take a set of situations S relative to which propositions about what you would do at home if you stay home is considered, but propositions about the consequences for your career are ignored. Within the set of situations in S compatible with your beliefs, it's *not* the case that in all of the situations where you stay home, your career is ruined. This is because within S, and so within the situations in S compatible with your beliefs, *there are no situations at all where your career is ruined* (nor are there situations where you career is not ruined). Rather, propositions about career ruin are undefined relative to S, since career ruin is ignored.

To see how the agent's desires come into play, it will help to lay out the semantics.

2.6 A situation semantics

How can you want a thing that you believe, all told, will result in something you greatly disprefer? That question poses a general problem for the semantics of 'want', which I've illustrated with a case study: a Best-worlds Semantics. In addressing this problem, I'll again use a case study: an adaptation of the Best-worlds Semantics, a Best-*situations* Semantics. The problem is general, and I intend for my solution to be too. Certainly, my solution can be implemented in semantic frameworks other than the Best-worlds framework. In the Conclusion, I discuss whether my solution can be implemented into all of the semantics that face our problem.

Two things change from the semantics that we began with—that is, the Best-worlds Semantics with the Belief Set Modal Base. First, situations will take the place of worlds. It is situations, rather than worlds, that are ordered by the ranking that the ordering source induces. And, crucially, the modal base contains situations, rather than worlds. Second, the modal base varies by context. This is because context-shift must be possible in our semantics: *Home*, recall, is true in some contexts and false in others. (We've been focusing on contexts where it's true, but of course in some contexts it's false; staying home would end in disaster!) What varies by context, I maintain, is what's considered and what's ignored, and that is represented by what's in a given set of situations; that is represented by what's in the modal base. In some

[17] A situation s is compatible with an agent A's beliefs just if there is no proposition p such that p is true in s and A believes not-p.

contexts, the modal base is a set of situations relative to which certain things are considered and other things ignored, while in other contexts, different things are considered and different things ignored, which is to say that the set of situations is different in that context. (More later on exactly how this delivers us both the context where *Home* is true and the one where it's false.)

The key carryover from before is that this new modal base, like the Belief Set Modal Base, is constrained by the agent's beliefs. In particular, what carries over is that the modal base never contains a possibility—that is, a world (as before) or a situation (now)—that is incompatible with the agent's beliefs. This is crucial—this is how the agent's beliefs don't play too small a role. For as we saw, when we evaluated (in Section 2.4) modal bases that contain worlds not compatible with the agent's beliefs, such worlds are irrelevant to the agent's desires.

To bring out this continuity between the Belief Set Modal Base and the new modal base, I state them together. Let S_c be the set of situations determined by c, a set relative to which certain propositions are considered and other ignored:

> *Situation Belief Modal Base (new)*
> $[\![\text{A wants p}]\!]^c$ is defined only if the modal base in c is the set of situations within S_c compatible with the A's beliefs.

The old modal base is this:

> *Belief Set Modal Base (old)*
> $[\![\text{A wants p}]\!]^c$ is defined only if the modal base in c is the set of worlds compatible with A's beliefs.

Just as when we introduced the Belief Set Modal Base, a caveat: for the Situation Belief Modal Base, I am setting aside cases where the agent either believes p or believes not-p, as well as cases where the advisory 'want' is at issue.[18]

Now the more technical details. As before, the ordering source determines an ordering $\preceq_{g_c(A)}$ that represents A's preferences. Everything is the same except it ranges over situations instead of worlds.[19] The modal base f_c is no longer a function from an agent to a set of worlds, but rather a function from an agent *and* the contextually determined set of situations S_c to a set of situations. In certain contexts, some things are considered and other things ignored, and this is represented by S_c: what's considered (ignored) relative to S_c is what's considered (ignored) in c. Analogously to before, $\text{best}_{A,f_c,g_c,S_c}$ comprises the best situations worlds in the modal base, $f_c(A, S_c)$, as ranked by $\preceq_{g_c(A)}$ (i.e. the agent's preferences).[20]

> *Best-situations Semantics*[21]

[18] See Grano and Phillips-Brown (ms) for a way to extend the Situation Belief Modal Base to cases where the agent either believes p or believes not-p.

[19] $s' \preceq_{g_c(A)} s''$ iff $\{p \in g_c(A) : p(s'') = 1\} \subseteq \{p \in g_c(A) : p(s') = 1\}$.

[20] $\text{best}_{A,f_c,g_c,S_c} = \{s' \in f_c(A, S_c) : \neg\exists s'' \in f_c(A, S_c), s'' \prec_{g_c(A)} s'\}$.

[21] Dandelet (ms) proposes a similar semantics. Crnič (2011) also sketches something simi-

$[\![\text{A wants p}]\!]^c = 1$ iff $\forall s \in \text{best}_{A,f_c,g_c,S_c}: p(s) = 1$.

Putting this semantics with the Situation Belief Modal Base gives the following new truth conditions.

> *New truth conditions*
> If defined, $[\![\text{A wants p}]\!]^c = 1$ iff, among the situations determined by c that are compatible with A's beliefs, p is true in the best of them, as ranked by A's preferences.

We are working up to a model of some-things-considered desire. In the last Section we modeled considering and ignoring, and how this fits with belief. With the semantics now on the table, we turn to the final piece of the puzzle: desire.

I claim that *Home* is true in a context (call it 'c_{Home}') where staying home is considered and its consequences for your career are ignored, meaning that c_{Home} determines a set of situations relative to which staying home is considered but its career consequences are ignored. The worlds within this set compatible with your beliefs make up the modal base in c_{Home}. You believe that if you stay home, you'll relax and paint; you believe that if you don't stay home, you'll do neither. And your beliefs about the career consequences of staying home versus not don't come into play, for those consequences are ignored. The modal base then looks like this:

modal base in c_{Home}	
stay-home situations:	*don't-stay-home situations:*
stay home	don't stay home
relax and paint	don't relax and don't paint
career is neither ruined nor not	career is neither ruined nor not

The situations where you you stay home are not awful, for the awfulness of the career consequences are ignored. Indeed, these situations are best, for in them you relax and paint, while in the other situations in the modal base, you neither relax nor paint. When staying home is considered but its career consequences ignored, *Home* is true.

2.7 What's considered and what's ignored?

On my view, certain things are considered, and others ignored. *Which* things are considered, and which ignored? I answered an instance of it in the last Section, explaining what's considered and what's ignored in the context where *Home* is true. In this Section, I'll answer other instances of the question about different contexts and cases.

lar, although his semantics works with questions, rather than situations, and is in response to Villalta's doxastic problem (see Section 8).

We've been concerned with cases where an agent wants something despite believing that it'll be accompanied by another thing that's truly awful for her. And we've seen what's considered and what's ignored in one such case. But what's considered and ignored in other cases of this kind? That is our first question. In all such cases, the wanted thing is considered, and the awful thing is ignored.

For example, we know that in the context where *Home* is true, the wanted thing, staying home, is considered, but the awful thing, the consequences for your career, is ignored. Why is this supposed to work in general? We have an agent A and contextually determined set of situations S, relative to which the wanted thing is considered and the awful thing ignored. Then, among the situations in S compatible with A's beliefs, there are situations where the wanted thing obtains, but no situations where the awful thing obtains (and no situation where it does not: the awful thing is ignored). The hypothesis is that these situations are best, and it would figure that they are. Without the awful thing dragging the wanted thing down, you prefer situations where the wanted thing obtains to those where it doesn't!

The second question returns us to our staying home case. While we've focused on finding a context where *Home* is true, there's also a context where it's false, as we've noted. (After all, staying home would ruin your career and that would be awful.) What's considered and what's ignored in the context where *Home* is false?

Home is false, I maintain, in a context where everything is considered and nothing is ignored. When the career consequences of staying home are considered, it's not true that you want to stay home, since the benefits of staying home are greatly outweighed by its badness of its career consequences.

More rigorously, the context where *Home* is false (call it '$c_{\neg Home}$') determines a set of situations relative to which every proposition is considered—i.e. the set of all situations relative to which every proposition is either true or false. This is simply the set all worlds, for as we noted when introducing the notion of a situation, a world just the limit case of a situation, one relative to which every proposition is either true or false. The modal base in $c_{\neg Home}$, then, is the subset of the set of all worlds that is compatible with the agent's beliefs, which is just her belief set! We have already encountered a semantics whose modal base is the belief set—it's the semantics of our case study, the Best-worlds Semantics with the Belief Set Modal Base. And with this semantics, as we know, *Home* is false. (Indeed, the fact that *Home* is false on this semantics is what presents the need for my view in the first place.)

There's a more general lesson here about what's considered and what's ignored. In any context where everything is considered and nothing is ignored, the modal base is the belief set. In any such context, *my semantics will make the same predictions as the semantics that we started with.* We get everything that was appealing about that semantics, and more besides. (For example, we can of course predict that there's a context where *Home* is true.)

I've said what's considered and what's ignored in the cases that have an-

imated this paper, but what about in other cases? This is our final question, and without an answer, we might fear overgeneration. If any old thing can be considered and any old thing ignored, we may predict predict readings for sentences that they do not have.

I do not have an algorithm for what's considered and what's ignored in any given case, and I'm skeptical that there is one. However, as far as the threat of overgeneration goes, I am on equal footing with two of the three rival views I discussed in Sections 3 and 4. (Of course I argued against these views on other grounds.) Take the view from Section 3, according to which ordering sources represent different values of the agent and vary by context. That view gave us a recipe for how to solve our problem cases (a recipe that ultimately wasn't general enough). But what about other cases? In any given context, what ordering source is selected? No proponent of a multiple ordering source view has given an answer to this question. Overgeneration looms. Similarly, consider the views (from Section 4) on which the modal base may contain worlds outside of the belief set. We saw what worlds these are in the problem cases, but what are they in other cases? Rubinstein (2017), according to whom the modal base sometimes contains circumstantially accessible worlds, provides no answer, leaving open the possibility of overgeneration. Anand and Hacquard (2013) say that every 'want' ascription is evaluated against a contrast class, but they don't provide a recipe to determine what that contrast class is in a given case.

2.8 Solving the doxastic problem

I identified a problem for the dominant semantics of 'want', which attempts to capture the relationship between believing and wanting. Villalta (2008) uncovered another, dubbed the 'doxastic problem' by Rubinstein (2017).[22] Recall that the idea behind the dominant approach to 'want' is that you're truly said to want p just in case you have a certain positive attitude towards what you believe will happen if p obtains. Now take some other proposition q, and imagine that what you believe will happen if q obtains is just the same as what you believe will happen if p obtains. You therefore have the relevant positive attitude towards q. In this case—when you believe that p if and only if q—you therefore want q. That is the problem. If you're truly said to want p and you believe that p if and only if q, it follows that you're truly said to want q. This inference is clearly invalid:

Instance of the doxastic problem

(i) You want to stay home.

[22]Three other potential solutions have so far been proposed. They come from Villalta (2008), who does not attempt to capture the relationship between believing and wanting; Rubinstein (2017), whom I objected to in Section 4; and Crnič (2011) who sketches, but does not develop, a question-based semantics similar to my situation-based one.

(ii) You believe that you will stay home if and only if your career is ruined.

(iii) $\Rightarrow$ You want your career to be ruined.

We can see concretely how the inference is validated within our familiar case study, the Best-worlds Semantics with the Belief Set Modal Base. The chain of reasoning is simple: by (i), stay-home worlds are best in your belief set; by (ii), the stay-home worlds are the career-ruin worlds in your belief set; so career-ruin worlds are best in your belief set, meaning that (iii) holds.

The inference pattern is invalid on my view if we adopt an entirely plausible constraint:[23] that to evaluate whether a desire ascription is true in a given context, its prejacent must not be ignored in that context.

Ignoring Constraint
$[\![\text{A wants p}]\!]^c$ is defined only if p is not ignored relative to S_c.

This constraint has intuitive appeal—it makes little sense to ask whether you want p when p is ignored—and theoretical considerations also recommend it. Without the constraint, $[\![\text{A wants p}]\!]^c = 0$ in any context where p is ignored, and we shouldn't be evaluating sentences in contexts where they are automatically false. $[\![\text{A wants p}]\!]^c = 1$ just if p is true in all of the situations worlds in S_c that are compatible with your beliefs. Yet if p ignored relative to S_c, then there are no situations at all within S_c where p is true. (This is just what it for a proposition to be ignored relative to a situation.) Compare the Ignoring Constraint to 'diversity constraints' for modals that ensure we don't evaluate sentences in contexts where they are automatically false or automatically true (Condoravdi, 2002).

We've already said that in the context where *Home* is true—i.e. the context where (i) is true—the proposition that your career is ruined is ignored. The Ignoring Constraint then dictates that (iii) is undefined in this context. The inference is blocked: (i) is true; so is (ii), we may imagine, but (iii) is not true but undefined.

2.9 The problem of logical entanglement

Before closing, let me address a problem, related to the one we've grappled with, that may have crossed your mind. In our cases, you want p despite the fact that you believe that p is entangled with something else you greatly disvalue. There are cases of a similar structure that have nothing to do with your beliefs at all: you want p despite the fact that p is, in a certain sense, *logically entangled* with something you greatly disvalue.

Imagine that you want to live a monastic life, dedicated to quiet contemplation and devoid of hedonistic pleasure. Yet much, much more important to

[23]The inference pattern is actually invalid even without the principle; but we should want the principle independently, as I think this paragraph brings out.

you is to live a life of hedonistic pleasure.[24] You want something to happen, to live a monastic life, despite the fact that if it does, something awful will by logically necessity happen as well: you won't live the (all-important) hedonistic life. (Leading a monastic life, as I'm imagining it, logically precludes living a hedonistic one.) Leading a monastic life and not leading a hedonistic life are logically entangled.

The some-things-considered desire approach does not help with this problem. That approach disentangles two things that are entangled within your beliefs via situations where the one thing is considered and the other thing is ignored. Staying home, for example, can be disentangled from career ruin with a situation where the former is considered and the latter ignored. Considering and ignoring can't disentangle logical entanglement. In any situation where you lead a monastic life, you do not live a hedonistic life, for living monastic life entails not living a hedonistic one. Leading a hedonistic life cannot be ignored relative to a situation where you lead a monastic life.

(A view where the modal base is not like the belief set, like the ones I considered in Section 4, also do not help with the problem: they disentangle, for example, staying home from career ruin by admitting into the modal base a world, outside of your belief set, where you stay home and don't have your career ruined. A monastic life and a hedonistic life can't be disentangled in this way. There are no worlds, within your belief set or without it, where you live a monastic life don't live a hedonistic one.)

While it might at first glance appear that the some-things-considered desire approach should be able to address logical entanglement, I believe that we should not expect it to. Cases of logical entanglement are of a fundamentally different kind than the cases that have animated this paper. As we've already seen, belief does not figure in cases of logical entanglement in the way it does in our cases. Further, desires behave differently with logical entanglement than in our cases.

Although a hedonistic life is all-important to you, you can truly be said *not* to want to live it. (You might say, 'My primary priority to is to live a hedonistic life, but to a lesser extent I also want not to live it, for I also want to live a monastic life instead.') In other words, you're truly said to want the greatly disvalued thing (not living a hedonistic life) that's logically entangled with what you want (living a monastic life). In the staying home case, by contrast, you cannot truly be said to want the greatly disvalued thing (having your career ruined) that's entangled by your beliefs with what you want (staying home).

This holds in general. In cases of logical entanglement, you want something despite it logically necessitating something else you greatly disvalue; in such cases, you can truly be said to want the greatly disvalued thing. A case adapted from Baker (2010) further illustrates the point. Imagine that you once

[24]Fara (2013, p. 268) forwards a similar case. Thank you to Kyle Blumberg for help developing this case.

watched a horror movie, but now never will again, for you greatly disvalue how frightened you felt when watching the movie. Yet in that fright, you also feel a thrill (this, you imagine, is why other people watch horror movies). You want to feel the thrill of watching a horror movie, but you intend never to feel that thrill again. We have logical entanglement: you want something (to feel the thrill of being frightened) that logically necessitates something you greatly disvalue (being frightened). Here again you can truly be said to want the greatly disvalued thing. (You might say, 'I want to be frightened by a horror movie just to feel the thrill, but of course the fright is ultimately unbearable so I won't do it.')

The difference in the behavior of desires in cases of logical entanglement versus our cases suggests the following distinction. In case of logical entanglement, you are in pieces, desideratively. There is a single outcome—feeling the fright of horror movies and the accompanying thrill—that you are split on, at once both wanting it and greatly disvaluing it. Not so in our cases. It's not *you* that's in pieces, but rather the object of your desires and values. There is no one thing that you at once greatly disvalue and want. Take the staying home case: what you greatly disvalue is the combination of the benefits of staying home and the consequences for your career—considering everything, it's not the case that your career to be ruined. In wanting to stay home, what you want is different: it's the benefits of staying home, where its career consequences are bracketed. Considering these benefits, but ignoring the career consequences, you want to stay home.

How might we model the idea that you are in pieces desideratively? I leave this question to another day.

2.10 Conclusion

On the dominant approach to the semantics of 'want'—motivated by the obvious connection between wanting and believing, and the theoretical fruits that capturing that connection brings—you're truly said to want p just if you have a certain positive attitude towards what you believe will happen if p obtains. This approach falters with case like the one where you want to stay home, despite having a negative attitude towards what you believe will happen if you do. I've proposed to view these cases differently, while preserving the connection between wanting and believing. You're truly said to want p just if you have a certain positive attitude towards what you believe will happen if p obtains—considering certain things and ignoring others. What's considered and ignored varies by context. In some contexts *Home* is true. You have a positive attitude towards what you believe will happen if you stay home—when painting and relaxing are considered, but career consequences are ignored.

I illustrated my solution by transforming a Best-worlds Semantics into a Best-situations Semantics. We can transform other semantics that take the dominant approach too. These semantics model an agent's beliefs in one of two ways: either with a set of worlds, such as the belief set, or a set of worlds

in combination with a credence function. Implementing my solution within the former approach is simple, and we've already seen it in action. We took the Best-worlds Semantics with the Belief Set Modal Base and replaced the belief set with a set of situations compatible with the agent's beliefs. The same can be done within any semantics that uses a set of worlds to model the agent's beliefs—e.g. Heim's (1992) or a more sophisticated variant of it, like one adapted from Blumberg (2018).[25] We simply swap in situations for worlds (and allow the relevant situations to vary by context). Things are more complicated for a semantics that uses both a set of worlds and a credence function (like Levinson's (2003) or mine in Chapter 1). Here too we substitute situations for worlds, but we must do more: modify the credence function to range over situations. Whether this can be done is a task for future work. It's a task worth taking on: credence-based accounts of 'want' face our problem, of course, and some-things-considered desire is a promising solution.

[25]Blumberg's account is for 'wish' but it can be simply changed to work for 'want' as well.

Chapter 3

Anankastic conditionals are still a mystery

3.1 Introduction

The *Harlem Sentence*, just below, doesn't look special.

(1) If you want to go to Harlem, you have to take the A train.[1]

It's clear what the sentence means, more or less, that taking the A train is necessary for going to Harlem. Yet a compositional account of its meaning, and the meaning of *anankastic conditionals* more generally, has proven an enigma.

Semanticists have assigned anankastics a unique status, developing accounts that distinguish them from ordinary indicative conditionals. Following Huitink (2008), Condoravdi and Lauer (2016) disagree, maintaining that, as their title says, "anankastic conditionals are just conditionals." I argue that Condoravdi and Lauer's account fails in the face of a well-known problem, *the problem of conflicting goals*. Their proposed solution to the problem works in certain cases: they rely on a special, *effective preference* interpretation for *want* on which what an agent wants cannot conflict with her *beliefs*. But a general solution requires that the goals cannot conflict with the *facts*. And so Condoravdi and Lauer's view doesn't work in all cases—indeed, I argue that it doesn't work in the most common cases.

In addition to proposing their semantics for anankastics, Condoravdi and Lauer introduce data of conditionals ('near-anankastics') that aren't anankastics but that nonetheless have the same compositionality problem as anankastics. The accounts previously given for anankastics, Condoravdi and Lauer argue (and I agree), do not generalize to near-anankastics. These accounts fail. And, I argue, so does Condoravdi and Lauer's. Anankastic conditionals are still a mystery.

I'll start by explaining the initial compositionality problem, what I'm calling the *problem of conditioning on goals*, and show how it's dissolved by Condoravdi and Lauer, who take their lead from Huitink. I'll then lay out the

[1]The name is from (von Fintel and Iatridou, 2005) and the sentence from (Sæbø, 2001).

problem of conflicting goals, Condoravdi and Lauer's proposed solution, and my argument against it. I'll consider, and ultimately reject, replies on Condoravdi and Lauer's behalf, as well as a different possible solution to the problem of conflicting goals.

3.2 The problem of conditioning on goals

Identified by Sæbø (1985, 2001), the problem of conditioning on goals is that the most straightforward application of Kratzer's (1981, 1991) classic system of modals and conditionals gets anankastics wrong.

In Kratzer's system, modals are evaluated against two conversational backgrounds, a modal base f and ordering source g, both functions from worlds to sets of propositions. Leaving the familiar details to a footnote, the semantics for necessity modals generally, and *have to* in particular, is:[2]

$$[\![\text{have to}]\!]^w(f)(g)(\lambda w.[\![\text{q}]\!]^w) = 1 \text{ iff } \forall w' \in \text{best}_{g(w)}(\bigcap f(w)): [\![\text{q}]\!]^{w'} = 1.$$

If modifies a modal base:

$$[\![\text{if r}]\!](f) = \lambda w.f(w) \cup \{\lambda w.[\![\text{r}]\!]^w\}.$$

Combining *if* and *have to* gives us:

$$[\![\text{have to}]\!]^w([\![\text{if r}]\!](f))(g)(\lambda w.[\![\text{q}]\!]^w) = 1 \text{ iff}$$
$$\forall w' \in \text{best}_{g(w)}(\bigcap(f(w) \cup \{\lambda w.[\![\text{r}]\!]^w\})): [\![\text{q}]\!]^{w'} = 1.$$

An anankastic conditional contains a *teleological* modal. Its ordering source provides certain relevant goals, its modal base certain relevant facts. Later, we'll consider what these goals and facts are. To see the problem of conditioning on goals, we can just stipulate the goals and facts.

The problem is this. Suppose that we're evaluating the Harlem Sentence in the actual world. Assume that throughout the modal base, various actually true propositions about New York's geography hold—that, for example, the A train is the only way to Harlem. Let the relevant goals be your actual goals, which, imagine, *don't* include going to Harlem. I assert the Harlem Sentence. The *if*-clause restricts the modal base to those worlds where you want to go to Harlem. We ask: do you take the A train in all of these worlds that best realize the relevant goals, that is, your *actual* goals? No—since your actual goals don't include going to Harlem! The Harlem Sentence comes out false even though the A train is the only way to Harlem.

A solution will say that when evaluating whether you have to take the A train in a world w, the proposition that you go to Harlem (*Harlem*) is a relevant goal in w.

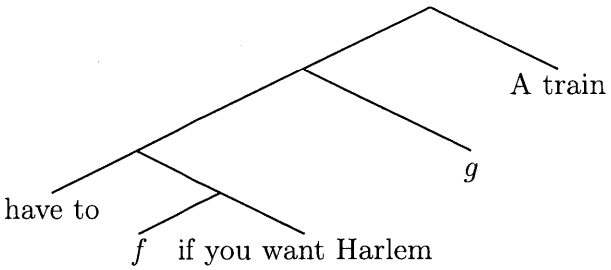
[2]We get a pre-order $\leq_{g(w)}$: $u \leq_{g(w)} v$ iff $\{p \in g(w) : p(v) = 1\} \subseteq \{p \in g(w) : p(u) = 1\}$. And where X is a set of worlds, $\text{best}_{g(w)}(X) = \{w'' \in X : \neg\exists v \in X[v <_{g(w)} u]\}$.

3.3 The first pieces of Condoravdi and Lauer's view

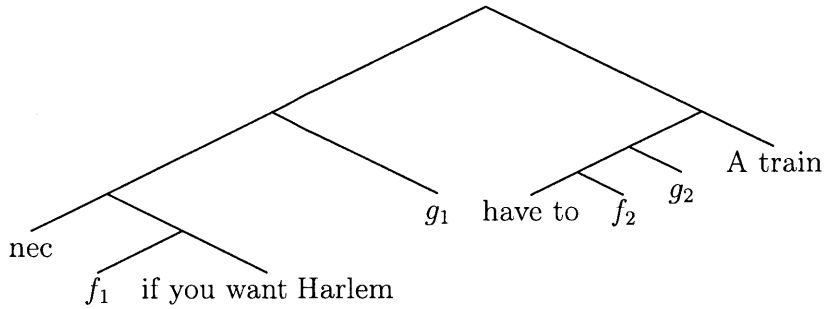
Authors have solved the problem of conflicting goals in various ways. Condoravdi and Lauer's solution belongs to a class of solutions that includes those of von Fintel and Iatridou (2006) and Huitink (2008), solutions on which the Harlem Sentence has a *double modal structure*. In addition to the overt modal, *have to*, there's a covert epistemic modal, *nec*, and it's *nec*, not *have to*, that's restricted by the *if*-clause. On Huitink's and Condoravdi and Lauer's views, anankastic conditionals are just ordinary indicative conditionals.

Here is the double modal structure, along with the single modal structure for contrast:

Single modal



Double modal



The semantics is:

$[\![\text{nec}]\!]^w([\![\text{if you want Harlem}]\!](f_1))(g_1)(\lambda w.[\![\text{have to}]\!]^w(f_2)(g_2)(\lambda w.[\![\text{A train}]\!]^w))$
$= 1$ iff $\forall w' \in \text{best}_{g_1(w)}(\bigcap(f_1(w) \cup \{\lambda w.[\![\text{you want Harlem}]\!]^w\}))$:
$[\![\text{have to}]\!]^{w'}(f_2)(g_2)(\lambda w.[\![\text{A train}]\!]^w) = 1$.

We evaluate the Harlem Sentence in a world w first by identifying a set of worlds (determined *nec's* conversational backgrounds, f_1 and g_1) where you want to go to Harlem. Then we ask whether you have to take the A train in each w' in the set. Advocates of the double modal view intend that (at least in typical cases) each world in *have to*'s modal base at w', $f_2(w')$, matches w in its subway facts. Suppose we're evaluating the Harlem Sentence in a world w where only the A train goes to Harlem. Then at each w' where we evaluate

whether you have to take the A train, every world in the modal base will be one where only the A train goes to Harlem. That looks right.

Condoravdi and Lauer's double modal view is their own because of their choices for the conversational backgrounds. We'll discuss the teleological ordering source later, since that's where my criticism lies. Consider the other three conversational backgrounds now. For reasons I won't get into, Condoravdi and Lauer say that *nec*'s modal base is *epistemic*, deriving from the speaker's true beliefs; *nec*'s ordering source is one of *typicality*; and *have to*'s modal base is *historical*.[3] So far, then, we have:

> *First pass semantics.* The Harlem Sentence is true in w iff
>
> a. For every most typical world w' compatible with the speaker's true beliefs in w *where you want to go to Harlem*:
> b. You have to take the A train in w'. More precisely:
> i. For every world w'' historically accessible from w' (each of which matches w in subway fact[4]) that best conforms to the relevant goals in w':
> ii. You take the A train in w''.

Note: in what follows, I will ignore the typicality constraint, since the cases I discuss can be filled out so that the constraint doesn't make a difference to my point.

The problem of conditioning on goals is solved if we require that wanting to go to Harlem in a world w' entails that going to Harlem is a relevant goal in w'. Then, since you want to go to Harlem in each w' where we evaluate whether you have to take the A train, going to Harlem is a relevant goal in w'.

The task is then to define the teleological ordering source in a way that entails this requirement—*while avoiding the problem of conflicting goals.* Condoravdi and Lauer's definition, which we'll see later, does entail the requirement;[5] they solve the problem of conditioning on goals. But, I argue, the problem of conflicting goals remains.

3.4 The problem of conflicting goals

The problem of conflicting goals, which must be faced by semantics of various kinds, manifests itself differently in different frameworks. I'll bring it out by showing how it falsifies a conjunction of two views: the first pass semantics, plus a first pass definition of the teleological ordering source, one that's often floated in the literature.[6]

Consider:

---

[3]See their pages 46 and 47.

[4]More precisely: each of these w'' matches w in subway fact when nothing atypical happens in w.

[5]To be precise, it entails a restricted version of this requirement. See page 49.

[6]See e.g., (Sæbø, 2001) and von Fintel and Iatridou (2005). (Sæbø, 2017) states it explicitly.

First pass definition of the teleological ordering source
If you want p in w, then $p \in g_2(w)$, the relevant goals in w.

This definition does rightly entail that if *Harlem* is wanted, then *Harlem* a goal. And it's natural to think that in evaluating what you have to do, what you want matters.

The problem is that we'll predict that the Harlem Sentence is false when it is intuitively true—in a case inspired by von Fintel and Iatridou's (2005) Hoboken Scenario. The prediction of falsehood follows purely from a supposition about what *the speaker's beliefs about the agent's desires*, and not from any facts about the subway.

The supposition couldn't be more common: that the speaker leaves open that you, the agent, *have two desires that can't both be realized.* More specifically, the speaker leaves open that you want to go to Harlem and want to do something else—say, go to Hoboken—that precludes going to Harlem. There is some world w' compatible with the speaker's beliefs where you both want to go to Harlem and to Hoboken, but can't go to both in w'. (The speaker needn't *believe* that the agent has two desires that can't both be satisfied, nor need the agent in fact have two desires that can't be satisfied; it's merely that the speaker's beliefs leave open that possibility.)

New Hoboken Scenario[7]

a. The A train is the only way to Harlem.
b. In some world w' compatible with the speaker's beliefs: you want in w' to go to both Harlem and Hoboken, but you can't go to both in w'.

I assert the Harlem Sentence. According to the first pass semantics, the sentence is false if there's a world w' compatible with my (the speaker's) true beliefs where you want to go to Harlem but do *not* have to take the A train. (Remember, we're ignoring the typicality constraint.) There is such a w'.

There's a world w' compatible with my beliefs—and thereby my true beliefs—where you want to go to both Harlem and Hoboken. The first pass definition of the teleological ordering source dictates that the relevant goals in w' include both *Harlem* and *Hoboken* (the proposition that you go to Hoboken). So, in some of the best worlds in the modal base at w' you go to Harlem, and in some you go to Hoboken. In none do you go to both, since you cannot go to both in w'. Supposing that the A train doesn't go to Hoboken in w', it follows that you do *not* take the A train in all of the best worlds in the modal base in w'. You do *not* have to take the A train in w'. The Harlem Sentence

[7]Von Fintel and Iatridou's original Hoboken Scenario is like the New Hoboken Scenario in that the A train is the only way to Harlem. It differs in that the speaker's beliefs about the compatibility of the agent's desire are not part of the scenario; rather, it's the compatibility of the agent's desires themselves. Von Fintel and Iatridou stipulate that you (the agent) in fact want to go to both Harlem and Hoboken, but in fact cannot go to both. (In footnote 11, I give a counterexample to Condoravdi and Lauer that closely resembles the original Hoboken Scenario.)

is predicted false.

But it is true. *Taking the A train is necessary for going to Harlem.* We've assumed nothing else that (it seems) should bear on the truth of the Harlem Sentence. (Recall that as a first approximation, the sentence means just this: that taking the A train is necessary for going to Harlem.) Our only other assumption is that it's compatible with my beliefs that you want two things that can't both be realized—in particular, going to Harlem and to Hoboken. And it couldn't be more normal leave this possibility open. All of us, all the time, want two things that can't both be realized.

Abstracting away from the particularities of the semantics here, the problem of conflicting goals at its core is this. In some world w' where we evaluate whether you have to take the A train, there are two goals in w', *Harlem* and *Hoboken*, that are jointly inconsistent with the facts in w'—there are two goals that *conflict with the facts.*

3.5 Condoravdi and Lauer's view in full

Recall that Condoravdi and Lauer and Huitink both solve the problem of conditioning on goals by positing a certain double modal structure. They also share a basic approach to the problem of conflicting goals, an approach on which *want* gets a special interpretation. They differ, though, on what that interpretation is. Condoravdi and Lauer's interpretation is situated in a new semantics for *want.*

I won't canvas the parts of their semantics not directly related to anankastics. We'll focus on their contention that *want* is sometimes interpreted against a special contextual parameter, *EP*, which represents what they call an agent's *effective preferences.* Wanting p in the effective preference sense—for short, wanting$_{EP}$ p—means that your desire for p is guiding your action. You might want to play in the NBA, but, knowing that's unattainable, your desire doesn't guide your action. You want to play in the NBA, but you don't want$_{EP}$ to play in the NBA. Or you might want to sleep, but want to go running more, and when you run, it's your latter desire that guides your action. Although you wanted to sleep, you didn't want$_{EP}$ to sleep. What you did want$_{EP}$ was to run.

Wanting$_{EP}$ is tightly linked to planning and intending. You want to play in the NBA, but you don't plan or intend to. You wanted to sleep, but it's running that you intended and planned to do. A close relative of wanting$_{EP}$, called *volitive wanting*, has long been discussed by philosophers (Davis (1984), from whom Condoravdi and Lauer take their cue, reviews the literature). Those who subscribe to the notion—and many do not—think that whatever you intend or plan to do, you can be truly said to want to do.

The crucial part of Condoravdi and Lauer's view is that you can't want$_{EP}$ two things that *conflict with your beliefs.*[8] More precisely: if you want *want*$_{EP}$

[8]Other semantics, including Heim (1992)'s and von Fintel (1999)'s, also disallow wanting

p and want$_{EP}$ q, you must believe that p and q can both be achieved—p and q must be jointly consistent with your beliefs.[9] This constraint is motivated in part by the idea that you can't be planning, or intending, to do two things that you believe cannot both be done (see Condoravdi and Lauer's pages 22–3 for more motivation).[10] For example, consider how strange it would be for someone to say:

(2) #I'm planning on going to Seattle tonight and I'm planning on going to Melbourne tonight, and I believe I can't do both.

Condoravdi and Lauer say that it's the effective-preference interpretation of *want* at play in anankastics:

(3) If you want$_{EP}$ to go to Harlem, you have to take the A train.

And the goals aren't merely what's wanted, as the first pass definition of the ordering source has it. Rather, they're what's wanted$_{EP}$:

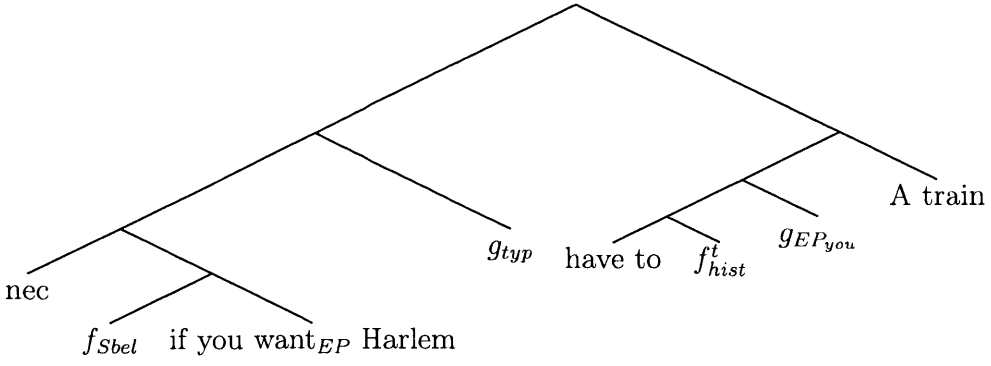
> *Condoravdi and Lauer's definition of the teleological ordering source*
> $p \in g_2(w)$ if and only if you $\underline{\text{want}_{EP}}$ p in w.

We can now lay out Condoravdi and Lauer's view in full. Adopting their nomenclature, f_{Sbel} is the speaker's-true-beliefs modal base, g_{typ} is the typicality ordering source, and f^t_{hist} is the historical modal base. The effective preference ordering source, $g_{EP_{you}}$, is this:

> $g_{EP_{you}}(w)$ contains exactly those propositions you want$_{EP}$ in w.

So we have:

Condoravdi and Lauer's logical form



> *Condoravdi and Lauer's semantics.* The Harlem Sentence is true in w iff

two things that conflict with your beliefs.

[9]This follows from Condoravdi and Lauer's stipulation that, in their terminology, the *preferential structure* that represents an agent's effective preferences in a given world obeys the *consistency* and *realism* constraints relative to her belief set (see their pages 29–31).

[10]This idea is a consequence of Grano's (2017) semantics for *intend*, which makes key use of effective preferences (see his pages 13–14).

a. For every most typical world w' compatible with the speaker's true beliefs in w where you $\underline{\text{want}_{EP}}$ to go to Harlem:
b. You have to take the A train in w'. More precisely:
   i. For every world w'' historically accessible from w' (each of which matches w in subway facts) that best conform to what you $\underline{\text{want}_{EP}}$ in w':
   ii. You take the A train in w''.

Focus on wanting$_{EP}$, since we've already reviewed everything else. Condoravdi and Lauer identify the goals with what's wanted$_{EP}$ in order to solve the problem of conflicting goals. After all, wanting$_{EP}$ is already conflict-free.

This identification works in certain cases. Take some world w' where we're evaluating whether you have to take the A train. Suppose that in w' you believe that it's impossible to go to both Harlem and Hoboken. Then, by the anti-conflict constraint on wanting$_{EP}$, you cannot both want$_{EP}$ to go to Harlem and want$_{EP}$ to go to Hoboken. Since the goals in w' are what you want$_{EP}$ in w', *Harlem* and *Hoboken* cannot both be among the goals! Your belief that *Harlem* and *Hoboken* conflict prevents the goals from containing two propositions that conflict with the facts.

3.6 The return of the problem of conflicting goals

To repeat: on Condoravdi and Lauer's view, the goals in a world w' won't contain both *Harlem* and *Hoboken* when you the agent believe that you can't get to both Harlem and Hoboken in w'. The problem is that you don't have this belief in every world. In some worlds, you believe that it's possible for you to go to both Hoboken and Harlem. Regardless of whether *Harlem* and *Hoboken* conflict with the facts in such worlds, the anti-conflict constraint on wanting$_{EP}$—which, to reiterate, bars conflict with your beliefs—doesn't kick in, meaning that you *can* want$_{EP}$ to go to both Harlem and Hoboken. *Harlem* and *Hoboken can* both be among the goals. The goals *can* conflict with the facts. That is the primary thesis of this paper.

We have the structural flaw in Condoravdi and Lauer's proposed solution to the problem of conflicting goals. Now consider a case, a modification of the New Hoboken Scenario, where the flaw is manifested. The Harlem Sentence is true, but it is predicted false. The prediction of falsehood follows entirely from a supposition about *the speaker's beliefs about the agent's plans (effective preferences) and beliefs*, and not from any fact about the subway.

As in the New Hoboken Scenario, the supposition could not be more common: the speaker leaves open that you, the agent, have two *plans* (two things you want$_{EP}$) that can't both be realized. More specifically, the speaker leaves open that you plan to go to Harlem and plan to do something else—say, go to Hoboken—that precludes going to Harlem. There is some world w' compatible with the speaker's beliefs where you plan (want$_{EP}$) to go both to Harlem and to Hoboken, but can't go to both in w'. (The speaker needn't *believe* that the

agent has two plans that can't both be carried out, nor need the agent in fact have two plans that can't both be carried out; it's merely that the speaker's beliefs leave open that possibility.)

Newer Hoboken Scenario

a. The A train is the only way to Harlem.
b. In some world w' compatible with the speaker's beliefs: you want$_{EP}$ in w' to go to both Harlem and Hoboken, but you can't go to both in w'.

We have this world w' where you want$_{EP}$ to go to both Harlem and Hoboken. Since the goals are supposed to be what's wanted$_{EP}$, the goals in w' include *Harlem* and *Hoboken*, which conflict with the facts in w'. (We are as before ignoring the typicality constraint.) As we know, when the goals conflict with the facts, the Harlem Sentence is predicted false.

But it is true. *Taking the A train is necessary for going to Harlem.* And, as before, we've assumed nothing else that (it seems) should bear whether the sentence is true. (Recall again that as a first approximation, the sentence means that taking the A train is necessary for going to Harlem.)

The Newer Hoboken Scenario is not a corner case. Its key assumption concerns the structure of the speaker's belief state, and that structure is shared by speakers in most cases where anankastics are asserted: the speaker leaves open the possibility that the agent has two plans (effective preferences) that cannot both be realized.

More specifically, the speaker leaves open that the agent simultaneously plans to achieve the hypothetical goal of the anankastic (e.g., going to Harlem) and plans to do to do something else—say, go to Hoboken—that precludes going to Harlem. As noted above, the speaker needn't believe that the agent has such plans. And, crucially, the speaker needn't have any particular plan in mind that could conflict with going to Harlem. The speaker need only leave open that the agents has *some plan or other* that could conflict with going to Harlem.

For a speaker not to leave open such a possibility would be for her to believe that the agent is so knowledgeable about the world—so good at accounting for every possible eventuality—that certain of her plans cannot conflict. In most cases, none of us are so knowledgeable. In most cases, speakers assume their addresses aren't so knowledgeable. In most cases, then, when a speaker asserts an anankastic, the structure of her belief state matches that of the speaker's in the Newer Hoboken Scenario.

To sum up: the goals may conflict when the agent's effective preferences conflict with the facts. And for Condoravdi and Lauer's view to fail, the agent's effective preferences needn't *actually* conflict with the facts. Rather, as will commonly be the case, the speaker just needs to leave open the possibility that they do.

3.7 Replies on Condoravdi and Lauer's behalf? (Looks like not)

Condoravdi and Lauer do recognize that identifying the goals with what's wanted$_{EP}$ does not guarantee that the goals don't conflict with the facts. They say two things about this, neither of which, I argue, will save their view.

First, they write:

> Generally and by default [the speaker] can assume that the agent is sufficiently informed about the relevant facts, so as to not have incompatible effective preferences (given the facts). (p. 49)

If this were indeed the default assumption, we should be suspicious of the Newer Hoboken Scenario. In it, I the speaker violate the assumption! I leave open the possibility that you the agent have incompatible effective preferences—in other words, that you want$_{EP}$ two things that conflict with the facts. A case that violates a default assumption is a case to be wary of.

This assumption is not the default, though, nor should it be. (There are also counterexamples, which I'll leave to a footnote, that don't violate the assumption.[11]) As I pointed out in the previous section, life with limited information leads us to have plans—effective preferences—that can't all be realized. Speakers know this: they leave open that their addressees have incompatible effective preferences.

Further, speakers can acknowledge that they leave open that their addressees have incompatible effective preferences. For example, I might say:

(4) If you want to go to Harlem, you have to take the A train. (But going to Harlem may mean that you'll be unable to do something else you may be planning to do. I can't know exactly what your plans are or just what might conflict with going to Harlem.)[12]

Here, the speaker acknowledges in the parenthetical that she leaves open that you may have a plan—an effective preference—that conflicts with going to Harlem. Not only is (4) fine to say, it seems to go without saying. (The first sentence of the parenthetical in particular seems seems so obvious that asserting it feels condescending.)

Condoravdi and Lauer's second concern is with cases that have *informational asymmetry* between the speaker and the agent. Consider the following (it's the second iteration of what Condoravdi and Lauer call 'the Virus Scenario' (p. 50)).

> *Virus Scenario* The A train is the only way to Harlem. Yet anyone who goes to Harlem will be infected by a virus that has entered the air

[11] For example: (i) the A train is actually the only way to Harlem; (ii) you actually want$_{EP}$ to go to both Harlem and Hoboken; (iii) you actually can't go to both; (iv) I *do* actually (and falsely) assume that you are sufficient informed about the relevant facts. I'll leave it to you the reader to work out how this is indeed a counterexample.

[12] Thank you to Magdalena Kaufmann for help coming up with this sentence.

there. You do not know about the virus, but I do. I'm not able to tell you about it right now.

a. (You:) How do I get to Harlem?
b. (Me:) You don't know all the facts, so don't do anything until I brief you in private, but if you want to go to Harlem, you have to take the A train.

Condoravdi and Lauer predict that the Harlem Sentence is false in this scenario.[13] (Without running through the derivation here, it's false because the anti-conflict constraint on wanting$_{EP}$ does not prevent conflict with the facts.) The case then appears to be a counterexample: the Harlem Sentence is predicted false in a world where the A train is the only way to Harlem.

There is something quite strange about (b), though, and the strangeness casts doubt on the import of the Virus Scenario—or so claim Condoravdi and Lauer. The reason for the strangeness of (b), they say, is the informational asymmetry between speaker and agent. Specifically, the informational asymmetry with respect to *Harlem* and *Not Infected*, the proposition that you won't be infected. We have that: *Harlem* and *Not Infected* conflict, I know of the conflict, and you don't. Condoravdi and Lauer suggest that in general, we cannot take at face value any apparent counterexample if there's informational asymmetry at play. I'm happy to agree. (Note, though, that to my ear and that of many informants, there's nothing strange about the Harlem Sentence in the Newer Hoboken Scenario, marking a dissimilarity between the Newer Hoboken Scenario and the Virus Scenario.)

So, *if* there were informational asymmetry in the Newer Hoboken Scenario, we should be suspicious of whether it is indeed a counterexample. (Without going into just why, Condoravdi and Lauer would say that the place to look for informational asymmetry would be with respect to *Harlem* and *Hoboken.*)

We're free to suppose that there is no informational asymmetry with respect to *Harlem* and *Hoboken* in the Newer Hoboken Scenario. For example, we can imagine that *Harlem* and *Hoboken* actually conflict with the facts, you don't know of the conflict, and neither do I.[14]

It's compatible with the original description of the Newer Hoboken Scenario that Harlem and Hoboken actually conflict because the description is silent on how Harlem and Hoboken actually relate. It's silent too on your state of mind, so there's no problem with the supposition that you don't know about the conflict. Finally, the description is also compatible with my not knowing

[13]More precisely, they predict that the Harlem Sentence is false *on its anankastic interpretation.* This is important for their discussion of the Virus Scenario, but we needn't trace out its implications here.

[14]Indeed, we can go further and suppose that as far as *Harlem* and *Hoboken* go, the speaker and agent have exactly the same knowledge and beliefs. For example, we may assume that the beliefs of the speaker and agent both leave open the possibility that *Harlem* and *Hoboken* conflict, and both leave open the possibility that they don't conflict. Further, we may imagine that speaker and agent have the exact same justification for leaving open these possibilities. All of this is compatible with the Newer Hoboken Scenario.

about the conflict. The only thing it says about me is that my beliefs leave open a possibility where the conflict exists but you nonetheless want$_{EP}$ to go to both. The Newer Hoboken Scenario stands as a counterexample.

3.8 A different solution? (Looks like not)

As we know, Condoravdi and Lauer address the problem of conflicting goals with a special interpretation for *want*. This might make you wonder: even though wanting$_{EP}$ can only prevent conflict with the agent's beliefs, is there a different interpretation that prevents conflict with the facts? An interpretation on which, for example, if you want to go to Harlem, you thereby don't want to do anything else that in fact conflicts with going to Harlem, regardless of your beliefs.

While the most common interpretation of *want* is intimately wrapped up with the agent's beliefs—as reflected in the literature on *want*[15]—there is another interpretation that's instead connected to the facts. To illustrate, take a case inspired by (Williams, 1981). Toni is about to drink from a bottle that she believes contains gin, but that in fact contains gasoline. I am aware of this. I say:

(5) Toni doesn't *really* want to drink from the bottle. (It contains gas!)

(6) (To Toni:) You don't *really* want to drink from the bottle. (It contains gas!)

With (5) and (6), Toni's *beliefs* aren't what's at issue: she believes that the bottle contains water, which she would enjoy. Rather, what matters are the *facts*: the bottle in fact contains gas, which she very much wouldn't enjoy. Intuitively, that's why (5) and (6) are true.

Following Jerzak (2019), call the interpretation of *want* in (5) and (6) the 'advisory *want*'.[16] The hope would be that (i) what you advisory-want cannot conflict with the facts, and (ii) the advisory *want* is the *want* of anankastics.[17] It's unclear whether the advisory *want* could prevent conflict with the facts, but even if it could, there are two problems.

First, speakers of French, Hindi, and Turkish report that their correlates of (5) and (6) are either false or infelicitous—and similarly for translations of English sentences that contain the advisory *want* more generally.[18] All of

[15]See e.g., (Heim, 1992) and (von Fintel, 1999).

[16]Jerzak extensively discusses the advisory *want*.

[17]A different approach from (ii) would be to say that want$_{EP}$ *is* the *want* of anankastics, as Condoravdi and Lauer claim, but that in problem cases like the Newer Hoboken Scenario, the Harlem Sentence is not interpreted as an anankastic, but rather as a conditional with the advisory *want*. The objections I raise below apply just as well to this approach.

[18]Here are French translations for (5) and (6), respectively:

(i) a. Toni ne veut pas *réellement* boire ce qu'il y a dans cette bouteille. (Elle contient de l'essence!)

b. Toni NE wants not *really* drink what that-there is in that bottle. (It contains

these speakers report that the Harlem Sentence is true in the Newer Hoboken Scenario. It would be surprising if in the Newer Hoboken Scenario the advisory *want* appears in the Harlem Sentence in languages in which it's otherwise absent.

Second, even in English, the advisory *want* can't help with all anankastics. Consider that conditionals that don't feature *want* at all can be interpreted as anankastics:[19]

(7) If you intend to go to Harlem, you have to take the A train.

(8) If you're planning on going to Harlem, you have to take the A train.

As with the Harlem Sentence (which does contain *want*), these anankastics raise the problem of conflicting goals. With the Harlem Sentence, the relevant goals were what's wanted, in some sense or other. What are the relevant goals for (7) and (8)? At a first pass: what you *intend* to do and are *planning* to do, respectively. But in the Newer Hoboken Scenario, there's a world compatible with my beliefs where you can't go to both Harlem and Hoboken and yet you intend to both and plan to go to both (recall that wanting$_{EP}$ is tightly connected with intending and planning). Again, the goals conflict with the facts.

We're seeing how far we can go with the idea that (i) it's the advisory *want*

---

some-of the-gas!)

c. 'Toni doesn't *really* want to drink from the bottle. (It contains gas!)'

(ii) a. Tu ne veux pas *réellement* boire ce qu'il y a dans cette bouteille. (Elle contient de l'essence!)

b. You NE want not *really* drink what that-there is in that bottle. (It contains some-of the-gas!)

c. 'You don't *really* want to drink from the bottle. (It contains gas!)'

The Turkish:

(iii) a. Toni aslında o şişeden içmeyi istemiyor. (İçinde gaz var!)

b. Toni actually that bottle-from drink want.NEG.IMPF. (Inside gas EXIST!)

c. 'Toni doesn't *really* want to drink from that bottle. (It contains gas!)'

(iv) a. Aslında o şişeden içmeyi istemiyorsun. (İçinde gaz var!)

b. Actually that bottle-from drink want.NEG.IMPF.2SG. (Inside gas EXIST!)

c. 'You don't *really* want to drink from the bottle. (It contains gas!)'

The Hindi:

(v) a. Toni us botal se *sac mein* piina nahi caahti hain. (us mein gas hain!)

b. Toni the bottle from *really* drink not want to. (in it gas there is!)

c. 'Toni doesn't *really* want to drink from that bottle. (It contains gas!)'

(vi) a. Tum us botal se *sac mein* piina nahi caahti ho. (us mein gas hain!)

b. You the bottle from *really* drink not want to. (in it gas there is!)

c. 'You don't *really* want to drink from the bottle. (It contains gas!)'

[19]Condoravdi and Lauer also affirm that (7) and (8) have anankastic readings (see e.g., their page 2), and they rely on the existence of *intend*- and *plan*-anankastics more generally to make various points (see their pages 20 and 23).

in anankastics (or at least anankastics with *want*) and (ii) you can't advisory-want two things that are incompatible with the facts. This would solve the problem of conflicting goals for the Harlem Sentence. But clearly it doesn't solve the problem of conflicting goals for (7) and (8). These sentences don't contain *want*! Further, there's no analogue of the advisory *want* that could apply to (7) or (8), no analogue that could resolve the problem of conflicting goals for anankastics with *intend* or *plan* more generally. Consider:

(9) #Toni doesn't *really* intend to drink from the bottle.

(10) #(To Toni:) You don't *really* intend to drink from the bottle.

(11) #Toni isn't *really* planning to drink from the bottle.

(12) #(To Toni:) You aren't *really* planning to drink from the bottle.

Even if the advisory *want* could help in certain cases, it can't furnish a general solution to the problem of conflicting goals.

3.9 Conclusion

Anankastic conditionals have been an enigma to semanticists, who've been led to distinguish anankastics from ordinary indicative conditionals. Condoravdi and Lauer, like Huitink, instead posit a double modal structure (thereby solving the problem of conditioning on goals) on which anankastic conditionals are just ordinary indicative conditionals.

I've argued that Condoravdi and Lauer don't have a general solution to the problem of conflicting goals. Their effective preference interpretation for *want* prevents conflict with the agent's beliefs. What we need, though, is to prevent conflict with the facts. When an agent is wrong about the facts, the goals may conflict. And *the agent needn't actually be wrong.* Condoravdi and Lauer's view fails whenever *the speaker leaves open that the agent is wrong* about certain facts (and leaves open that she has certain plans), as the speaker will in most cases.

Where does that leave us? Not somewhere good. Remember that Condoravdi and Lauer argued that previous semantics for anankastics don't generalize to near-anankastics. If Condoravdi and Lauer are right about that, and I believe that they are, then those semantics are inadequate. But so is Condoravdi and Lauer's. Although I wish that I could point to a way forward, I can't see one. Anankastic conditionals are still a mystery.

Chapter 4

Getting what you want (with Lyndal Grant)

4.1 Introduction

A widely shared sentiment, articulated by Dennis Stampe, is that desire satisfaction is 'truth by a different name' (1986, p. 154). The sentiment can be sharpened by appeal to two principles, one about belief and the other about desire:

> *Truth-is-Truth Principle*
> If A believes p, then A has a belief that is true in exactly the worlds where p is true.[1]
>
> *Satisfaction-is-Truth Principle*
> If A wants p, then A has a desire that is satisfied in exactly the worlds where p is true.[2]

The Truth-is-Truth Principle is true. But, we will argue, the Satisfaction-is-Truth Principle is not. An agent may want p without having a desire that is satisfied in exactly the worlds where p is true—in particular, without having a desire that is satisfied in every world where p is true. Such an agent has a desire whose satisfaction conditions are what we call *ways-specific*: it is satisfied only when p obtains in certain ways.

(The Satisfaction-is-Truth Principle presupposes that desire is a propositional attitude.[3] Whether this presupposition is true is orthogonal to our

[1] Though widely accepted, Bach (1997) questions a principle in this vein.

[2] See e.g. (Searle, 1983, ch. 2), (Whyte, 1991), (Stampe, 1994), (Heathwood, 2006). Condoravdi and Lauer (2016) give a contextualist take on the principle. Braun (2015) endorses a similar principle, which he calls 'The Weak Content-Specification Version of the Relational Analysis of Desire Ascriptions' (on which more in §4.10): 'If N is a proper name and S is an infinitival phrase (with or without explicit subject), then: if $\ulcorner N$ wants $S\urcorner$ is true, then the referent of N has a desire that is satisfied in exactly those worlds in which the proposition that S semantically expresses is true' (p. 149).

[3] A presupposition contested by e.g. Montague (2007) and Moltmann (2013).

argument, which works just as well against a version of the Satisfaction-is-Truth Principle that doesn't mention propositions: if A wants to φ, then A has a desire that is satisfied in exactly the worlds where she φs.)

Consider a case. Millie says that she wants to drink milk. Suppose (and we'll revisit this supposition later) that she is right. Intuitively, Millie nonetheless does not have a desire that is satisfied when she drinks spoiled milk. Millie wants to drink milk, but, intuitively, not just any old milk will do.

To show that a case like Millie's is a counterexample to the Satisfaction-is-Truth Principle, we need to establish two claims. First, agents like Millie do want what they say they want—e.g. Millie does want to drink milk. Second, Millie indeed does not have a desire that is satisfied when she drinks spoiled milk, and similarly for agents like her.

Fara (2003, 2013) and Lycan (2012, ms) accept similar claims on the basis of similar cases.[4] We provide new arguments for both claims. Our arguments for the first go beyond those offered by Fara and Lycan for analogues of our first claim.

The only support they offer for claims analogous to our second claim is intuitions about when agents get what they want—e.g the intuition that Millie doesn't have a desire that is satisfied when she drinks spoiled milk.[5] As you might expect, these intuitions have been contested (by Braun (2015) and Prinz (ms), as cited in (Lycan, 2012, pp. 205–6)). These contested intuitions about getting what you want play no role in our argument. Instead, we argue by appeal to the *dispositional role* of desire. Because agents are disposed to satisfy their desires, an agent's dispositions provide important evidence about the satisfaction conditions of her desires. That evidence, we argue, shows that desire satisfaction is indeed ways-specific.

4.2 The argument

Here is our argument at a high level: agents are disposed to satisfy their desires; desire-based dispositions are ways-specific; so, desire satisfaction is ways-specific.

To begin, let's fill out the case of Millie and the spoiled milk. Millie is eating a chocolate chip cookie, and says out loud to no one in particular, 'I

[4]Fara (2013) rejects a principle closely related to the Satisfaction-is-Truth Principle, which she calls the 'content-specification version of the relational analysis' (p. 254) of desire ascriptions. She gives only an instance of the principle: ' "Lora wants to be in London" is true just in case Lora has a desire that is satisfied in exactly those possible worlds in which she is in London' (p. 254) (in her (2003), she rejects a similar principle). The left-to-right direction of the principle—the direction that she objects to—is an instance of the Satisfaction-is-Truth Principle if we accept, as we should, that if Lora wants to be in London, then 'Lora wants to be in London' is true. See more in §4.10. Lycan isn't explicit about just what principles he objects to. We read him (2012, pp. 206–7; ms, pp. 2–3) as committed to the possibility of cases that would falsify the Satisfaction-is-Truth Principle. And in his (ms), he cites Fara's (2013) and seems to side with her (pp. 2–3).

[5]van Rooij (1999) and Persson (2005, ch. 10) also discuss these intuitions.

want to drink some milk, but the milk in the refrigerator is spoiled.' Although her path to the refrigerator is clear, Millie does not drink the spoiled milk. We'd like to suppose that Millie really does want to drink milk, and that she is not disposed to drink the spoiled milk. In §4.3–4.5, we'll discuss whether these are legitimate suppositions—whether the case as we suppose it to be really is possible. For now, we'll assume that the suppositions are legitimate: Millie wants to drink milk and she is not disposed to drink the spoiled milk.

Millie wants to drink milk, but she isn't disposed to drink the spoiled milk—she isn't disposed to drink the only milk that she believes is available to her. It's not that she isn't disposed to drink any kind of milk at all. She is. It's rather that her disposition to drink milk is discriminating. It is *specific* to certain kinds of milk. Not just any old milk will do.

Millie has what we call a *ways-specific desire-based disposition.* If an agent has a ways-specific desire-based disposition, then for some p, (i) she wants p; (ii) there are ways for p to obtain that she is disposed to bring about; but (iii) there are other ways for p to obtain that she is not disposed to bring about, even if she believes that she can only bring it about that p obtains in those ways. Because Millie's disposition is specific to certain ways of its being the case that she drinks milk—ways in which she drinks certain kinds of milk—it is ways-specific in just this sense.

To run our argument, we need to state carefully the thesis that agents are disposed to satisfy their desires. Here's how others have stated the thesis:

> [T]he primitive sign of having a desire is trying to satisfy it. (Humberstone (1990, p. 107), riffing on Anscombe)
>
> [T]he actions a desire is a disposition to perform are those that would satisfy that desire provided the agent's operative beliefs were true. (Stampe, 1994, p. 246)
>
> [A] desire is manifested in... behaviour aimed at satisfying the desire. (Hyman, 2014, p. 85)

In stating the thesis ourselves, we commit only minimally on further questions concerning how desires relate to dispositions. We do not assume, for example, that desires *are* dispositions. And, as far as we're concerned, the principle can be contingent, or restricted to certain kinds of agents.[6] We propose:

> *Satisfaction–Disposition Principle*
> If A has a desire that is satisfied in exactly the worlds where p is true, then A is disposed to do what she believes will bring it about that p obtains.[7]

[6]It needn't apply, for example, to agents incapable of action, like Strawson (1994, ch. 10)'s "Weather watchers."

[7]A weaker version of this principle that employs an 'other things equal' clause to accommodate troublesome cases would work just as well for our purposes, as we explain in §4.6.

Now the argument.

P1. If Millie has a desire that is satisfied in exactly the worlds where she drinks milk, then Millie is disposed to do what she believes will bring it about that she drinks milk. (instance of the Satisfaction–Disposition Principle)

P2. Millie wants to drink milk.

P3. Millie is not disposed to do what she believes will bring it about that she drinks milk—she is not disposed to drink the spoiled milk .

C1. Millie does not have a desire that is satisfied in exactly the worlds where she drinks milk. (by P1 and P3)

C2. Millie wants to drink milk and Millie does not have a desire that is satisfied in exactly the worlds where she drinks milk. (by P2 and C1)

C2 is a counterexample to the Satisfaction-is-Truth Principle, which entails that if Millie wants to drink milk, then she has a desire that is satisfied in exactly the worlds where she drinks milk.[8]

In its basic form, our argument then is this: agents are disposed to satisfy their desires (P1); desire-based dispositions are ways-specific (P2 and P3); so, desire satisfaction is ways-specific (C2).

Now we'll defend the premises.

4.3 In defense of P2: on saying something false but helpful

In defending the premises, we claim first that *a certain principle is true*—the Satisfaction–Disposition Principle (P1). We claim second that *a certain kind of case is possible*—one where Millie wants to drink milk (P2) and isn't disposed to drink the spoiled milk, despite believing it's the only milk available to her (P3).

In arguing for P2 and P3, then, we are arguing for the possibility that P2 and P3 are true together. In this section and the next, we are concerned with defending P2. We'll assume that P3 is true and maintain that it's possible for P2 to be true as well. In §4.5, we'll assume that P2 is true and maintain that it's possible for P3 to be true as well.

[8] The Satisfaction-is-Truth Principle says that if A wants p, then A has a desire that is satisfied in exactly the worlds where p is true. So, strictly speaking, C2 is a counterexample to the Satisfaction-is-Truth Principle just in case the proposition denoted by the complement of 'want' in 'Millie wants to drink milk' is one that's true in exactly the worlds where Millie drinks milk (for more see §4.10). Of course it seems to be such a proposition that's denoted! (It is not, for example, the proposition that Millie drinks milk or stubs her toe.) You might worry, though, that in fact it's a different proposition. We defer here to Fara (2013), who argues extensively that the complements of desire ascriptions like 'Millie wants to drink milk' do denote the propositions that they seem to.

Turn now to the argument for P2. Millie, recall, asserts that she wants to drink milk. Suppose that Millie speaks sincerely and is as good as anyone at knowing what she wants. The default position here should be that Millie does want to drink milk. That is, after all, how things would seem if you were faced with someone like Millie, who gives a sincere, well-informed report of what she wants.

(To be totally clear: in maintaining that it's true that Millie wants to drink milk, we don't mean to implicate that it isn't also true that Millie wants to drink *fresh* milk. Indeed, we think it's both true that Millie wants to drink milk and true that Millie wants to drink fresh milk!)

An imaginary interlocutor might resist our claim that it's possible that Millie wants to drink milk (while not being disposed to drink the spoiled milk). The interlocutor would then need a hypothesis about why it's so natural to think that Millie does want to drink milk. Below is one such hypothesis; in the next section we consider another.

Often we say things that are false because a falsehood is most helpful for what we're trying to communicate (see e.g. (Lasersohn, 1999)). Take a case adapted from (Sperber and Wilson, 1985). Brigitte lives in Issy-les-Moulineaux, which is just outside the city limits of Paris. At a party in London, Brigitte is asked where she lives. She replies:

(1) [Brigitte:] I live in Paris.

(1) is false, since Brigitte lives just outside the city limits of Paris. Nonetheless, (1) serves its communicative purpose perfectly well.

The hypothesis is that when Millie asserts (2) she is just like Brigitte: she says something false but helpful.

(2) [Millie:] I want to drink milk.

Millie is *un*like Brigitte though. Here's why.

Brigitte must *retract* (1) in the face of the truth. Suppose that you hear Brigitte and say:

(3) [You:] Actually, Brigitte doesn't live in Paris. (She in fact lives in Issy-les-Moulineaux, which is outside of Paris.)

If Brigitte is pressed—which is it, in Paris, or just outside the city limits?—she'd be under pressure to retract:

(4) [Brigitte:] You are right; I don't live in Paris.

Brigitte must retract her original statement because one can't both live in Paris and outside of Paris (assuming one lives in just one place).[9]

But Millie does not need to retract (2) under pressure.[10] Suppose that you hear Millie and say:

[9]Yablo (2014, ch. 5) makes a similar point.

[10]Thank you to Kai von Fintel (p.c.) for this point.

(5) [You:] Actually, Millie doesn't want to drink milk. (She in fact wants to drink fresh milk.)

If Millie is pressed—which is it, milk, or fresh milk?—she isn't under pressure to retract. She does *not* have to say:

(6) [Millie:] You are right; I don't want to drink milk.

While it can't both be true that one lives in Paris and true that one lives outside of Paris, it *can* both be true that one wants to drink milk and true that one wants to drink fresh milk. And, again, that is exactly what we say about Millie: it's true that she wants to drink milk, and it's true that she wants to drink fresh milk.

We can further bring out the dissimilarity between Millie's and Brigitte's cases by considering a third case, one in which the speaker says nothing false. Suppose that Yannick lives in the Marais, which *is* in Paris. At a party in London, Yannick is asked where he lives.

(7) [Yannick:] I live in Paris.

Suppose that you hear Yannick and say:

(8) [You:] Actually, Yannick doesn't live in Paris. (He in fact lives in the Marais, which is in Paris.)

This is nonsense! Yannick is under no pressure at all to retract (7). It's true that he lives in the Marais *and* it's true that he lives in Paris. Yes, Yannick could give you more information about where he lives by saying (9) instead of (7):

(9) [Yannick:] I live in the Marais.

But just because the one statement is more informative than the other does not make the first false.

The same goes for Millie. Yes, she could give you more information about what she wants by saying (10) instead of (2):

(10) [Millie:] I want to drink fresh milk.

But, again, just because the one statement is more informative than the other doesn't make the first false.

To summarize. Brigitte says one false but helpful thing (she lives in Paris) and one true thing (she lives just outside of Paris). Yannick says *two* true things, one of them (he lives in Paris) less informative than the other (he lives in the Marais). We say that Millie is more like Yannick than like Brigitte: Millie says two true things, one of them (she wants to drink milk) less informative than the other (she wants to drink fresh milk).

The analogy between Yannick and Millie is imperfect. While living in the Marais entails living in Paris, it's controversial whether wanting to drink fresh milk entails wanting to drink milk.[11] However, our point remains: saying that

[11]Heim (1992), for example, says that it doesn't, while von Fintel (1999) says that it does

Millie wants to drink milk doesn't specify everything about what she wants, just as saying that Yannick lives in Paris doesn't specify everything about where he lives. It's nonetheless true that Yannick lives in Paris. Likewise, we claim, it's nonetheless true that Millie wants to drink milk. A desire report need not be maximally specific in order to be true. Millie doesn't fully specify what she wants, but nevertheless what she says is true.

The dialectic in this section has been this. Supposing that Millie is not disposed to drink the spoiled milk, we've argued that it's possible that P2 is true—that Millie wants to drink milk. Our imaginary interlocutor contested this, hypothesizing that it must be that Millie said something false but helpful. As we've seen, though, this hypothesis fails.[12]

Millie's case could of course be filled out so that she does not want to drink milk. But it clearly makes sense, and in fact seems most natural, to take Millie at her word.

4.4 In defense of P2: on saying and asserting

In this section we consider a different hypothesis about why it's so natural to think that Millie wants to drink milk even if, as our imaginary interlocutor argues, Millie doesn't in fact want to. This hypothesis co-opts a distinction made by Braun (2015) between what one *says* and what one *asserts*.

According to Braun, you can say a certain proposition while at the very same time asserting various other propositions. Suppose you say p and p is false. When you say p, you may at the very same time be asserting some other proposition that is true. In such a case you said something false while asserting something true. In Braun's terminology, you have *spoken truly while saying something false* (see e.g. his p. 157).[13]

If Braun is right, then the following case is possible. Millie does not want to drink milk but says that she does. When saying that she wants to drink milk, she asserts some other proposition that is true—say, the true proposition that she wants to drink fresh milk. Our imaginary interlocutor could hypothesize that *this* is why it's so natural to think that Millie says something true when she says that she wants to drink milk, even if she does not in fact want to.

There are two ways resist this thought. The first would be to deny Braun's distinction between saying and asserting. Some may deny this, but we won't try to adjudicate the issue here.

The second way is to grant Braun's distinction, but resist our imaginary interlocutor's hypothesis. This is what we'll do, maintaining that Millie's case as we've described it is unlike the kind of case that Braun cites as a 'plausible

(see more in footnote 28).

[12]As we noted in the introduction, Fara (2003, 2013) and Lycan (2012, ms) also argue that seemingly true desire ascriptions, like (2), are indeed true.

[13]As precedents for his view, Braun cites similar distinctions made by Bach (1994, 2001, 2005) on saying and implic-*i*-ing; Soames (2005, 2008) on semantic content and asserting; and Braun (2011) on locuting and asserting.

example' (p. 157) of an agent using a desire ascription to assert something true while saying something false.[14]

Braun gives the following example (p. 157):

(11) [Suppose that Sara is teaching a philosophy seminar and suppose she has noticed that many of her students in her seminar arrived late. So she utters:] I want everyone to arrive on time for the next meeting of this seminar.

Braun invites us to suppose, following Bach (2000) and Soames (2005, 2008), that 'everyone' is never contextually restricted, that it always quantifies over all people in the universe. According to Braun, what Sarah *says* is the proposition that she wants every human in the universe to arrive on time for the next seminar meeting, but she *asserts* all at once various other propositions—among them the true proposition 'that Sarah wants everyone *to whom she is speaking* to arrive on time for the next meeting' (p. 158; emphasis in the original).[15] What she says is false (she does not want every human in the universe to arrive on time for the next meeting), but she nevertheless asserts a true proposition.

On our interlocutor's hypothesis, Millie is like Sarah. When Millie's dispositions are as we have supposed and she says that she wants to drink milk, she says something false but nonetheless asserts a true proposition, the proposition (say) that she wants to drink fresh milk.

But Millie is unlike Sarah, and retraction data again provide key evidence. Consider that if you insisted that Sarah doesn't really want *everyone* to come, she would be under pressure to retract, to disavow the proposition that she said. Take the following exchange, for example:

(12) [You:] Sarah doesn't want *everyone* to come to the next meeting on time! She just wants those to whom she was speaking to come to the next meeting on time!

(13) [Sarah:] Okay, fine. I don't want *everyone* to come; I just want those to whom I was speaking to come.

But as we saw in the last section, if you insisted that Millie doesn't really want to drink *milk*, she wouldn't be under pressure to retract.[16]

To summarize: we've claimed that it's possible that P2 is true—that Millie

[14]We should emphasize that Braun is *not* committed to saying that Millie's case, as we've described it here in §4.4, is like his plausible example. More generally, we are not objecting to Braun's views about language: we neither object to his saying–asserting distinction (as we noted), nor do we object to the argument in which he puts that distinction to use. Rather, what we object to is the argument of an imaginary interlocutor who co-opts Braun's distinction. (See more in footnote 16 on the relationship between Braun's argument and our own.)

[15]This is a slight simplification. Braun suggests that Sarah may say more than one proposition in uttering (11).

[16]Now, if we were to stipulate that Millie does not want to drink milk—Braun makes such a stipulation in an analogous case in his §8.1—then she *should* be under pressure to retract. But that is not what's stipulated here in §4.4; rather, it's what's at issue.

wants to drink milk, while assuming that she is not disposed to drink the spoiled milk. Our imaginary interlocutor contested this possibility, claiming that Millie said something false while nonetheless asserting something true. And while we may be able to imagine a version of our case in which this is in fact so, our interlocutor is committed to saying that if Millie is not disposed to drink the spoiled milk, she *must* be saying something false. This is what we deny.

4.5 In defense of P3: against the other desires hypothesis

Now P3: Millie is not disposed to drink the spoiled milk. In this section, we assume that P2—Millie wants to drink milk—is true, and argue that it's *possible* that P3 is also true. Suppose that you wanted to deny this possibility. Your claim would be that, given that Millie wants to drink milk, it *must* be that Millie is disposed to drink the spoiled milk. You'd then need a hypothesis about why Millie doesn't drink the spoiled milk, despite being disposed to drink it.

Here is such a hypothesis.

Start with something that everyone should agree on. How an agent acts depends not just on whether she has a certain desire and associated disposition, but also on what else she wants.[17] For example, suppose that Portia wants to buy a Porsche, and that she is disposed to buy a Porsche. She doesn't buy one, though, and that's because in addition to wanting to buy a Porsche, there's something else she wants: not to spend so much money that she is financially ruined. Her disposition to buy a Porsche isn't manifested because she wants this other thing.

According to the *other desires hypothesis* of Millie's inaction, Millie is like Portia. The hypothesis has two parts: (i) Millie *is* disposed to drink the spoiled milk, but (ii) she wants other things, preventing her disposition from manifesting.

Let's grant that Millie does want other things that bear on drinking the spoiled milk—e.g. she wants not to drink something sour, and she wants not to be sick to her stomach. The question is then whether her wanting these other things is interfering with the manifestation of a disposition to drink the spoiled milk—as the other desires hypothesis says. We think Millie has no such disposition.

To see why, contrast Millie with Portia, who, in being disposed to buy a Porsche, sees something in buying it: driving fast and making her friends envious. It makes sense that Portia would have a disposition to buy a Porsche—even though the disposition doesn't manifest itself—because a Porsche is alluring to her. But Millie sees nothing appealing at all in drinking the spoiled milk. What would the appeal even be? Everything that is normally appealing to Millie about milk is absent in the spoiled milk. Millie enjoys the mild flavor

[17] Ashwell (2017) develops a theory on the interactions among desire-based dispositions.

and smell of fresh milk; the spoiled milk is overpoweringly sour. Millie likes the smooth mouth feel of fresh milk; in the spoiled milk, the protein has separated from the whey, forming unpleasant clumps. Spoiled, separated milk doesn't even have the nice creamy look of fresh milk. Given that the spoiled milk has no appeal for Millie, why *would* she be disposed to drink it?

Even if you're not convinced by our argument against the other desires hypothesis in Millie's case, there are other cases relevantly like Millie's where the other desires hypothesis clearly fails. In these cases, the agent does not want any other things that could explain her inaction.

Consider Trina, whose neighbor has, much to Trina's dismay, just installed a full-scale plastic replica of Michelangelo's David. The sculpture is all too visible from Trina's kitchen window, and her view of it needs to be blocked tonight. Having a tree planted in between the sculpture and the window seems best: Trina wants to have a tree planted in her backyard by the end of the day. It so happens that Trina believes that the only trees available to her today are bonsais, which are too small to block her view of anything. Further, bonsais don't have the majestic look that Trina has always admired in trees of the size that could block the statue. Nothing that appeals to Trina about having a tree planted is present with a bonsai. The day ends without Trina trying to have a bonsai planted.

The other desires hypothesis would say that (i) Trina is disposed to have a bonsai planted, but (ii) she wants other things, preventing this disposition from manifesting.

But we can easily suppose that Trina doesn't want any such things. Imagine that you go to Trina's backyard with a bonsai in hand, dig up a few inches of dirt, and tell Trina that you might plant the bonsai—how does she feel about it? Trina says that she doesn't care. As we know, nothing appeals to her about the bonsai. But neither is there anything unappealing. Having it planted comes at no cost to her. You are proposing to plant it for her, so she wouldn't have to get her hands dirty. And you wouldn't put the bonsai in a place that would stop Trina from planting a tree that could block the statue. Nor would you plant it in a place that would impede the route that she normally takes when she walks across her yard, or...Even if Trina did want not to get her hands dirty or to have her normal route unimpeded, her desires would have no impact on whether she has a bonsai planted.

As far as Trina is concerned, it's fine if the bonsai is planted, and fine if not. Trina is *indifferent*. There's nothing she wants either way about the bonsai. In particular, there's nothing that she wants about the bonsai that would prevent the manifestation of a disposition to plant a bonsai. This contradicts the other desires hypothesis.

Consider Portia for contrast again. Portia is *ambivalent*. She is at once both attracted to buying a Porsche (it would mean fast driving and envious friends) and repelled by it (she'd surely go bankrupt). The unappealing features of buying a Porsche overwhelm the attraction, which is why Portia does not buy a Porsche. The other desires hypothesis makes perfect sense of the situation.

Given that Portia is both attracted to and repelled by the prospect of buying a Porsche, it's natural to think that she is both disposed *to* buy it, and that she wants other things that speak in favor of *not* buying it—things that prevent the disposition to buy it from manifesting. Not so with Trina. She is indifferent, neither attracted to nor repelled by the prospect of having a bonsai planted. It is her indifference that explains her inaction.

The other desires hypothesis fails with Trina. The point of the hypothesis is to explain why an agent does not act despite having a (hypothesized) disposition to act. No doubt Trina's case could be filled out so that Trina is disposed to have a bonsai planted, yet does not do so for some reason or other. But it clearly makes sense to fill it out in the way we have. If you want to maintain that Trina *must* be disposed to have a bonsai planted, you can't merely give a way of filling out the case so that Trina has an unmanifested disposition to have a bonsai planted; you must show that there is no possible way of filling it out as we have just done.

If you prefer Trina's case to Millie's, run our argument with Trina. Either way, P3 stands: the agent (Millie, Trina) is not disposed (to drink the spoiled milk, to have a bonsai planted).

4.6 In defense of the Satisfaction–Disposition Principle: on an 'other-things-equal' clause

The final premise of our argument to defend is P1, which is an instance of the Satisfaction–Disposition Principle. We'll dispel one potential worry about the principle in this section and then others in §4.7 and §4.8.

When in a bold mood, philosophers state connections between desires and dispositions in the same form that we've stated the Satisfaction–Disposition Principle: if an agent is in such and such a desire state, then she is disposed to act thus-and-so-ly, given certain beliefs. When in a cautious mood, philosophers add an 'other things equal' clause: if an agent is in such and such a desire state, then, *other things equal*, she is disposed to act thus-and-so-ly, given certain beliefs.

You might worry that Millie's case calls for a cautious mood—that it calls for a version of the Satisfaction–Disposition Principle with an 'other things equal' clause. If things were *un*equal with Millie, then our argument wouldn't go through.

Consider some ways for things to be unequal—ways for you to lack a disposition to do what you believe will satisfy your desire. You might be unaware of your desire, or have false second-order beliefs about your first-order beliefs about how to bring it about that your desire is satisfied, or be simply unable to bring it about that your desire is satisfied.

We can simply suppose that things are *not* unequal for Millie in these ways—that she is aware of her desires, that she believes that she believes that drinking the spoiled milk will bring it about that she drinks milk, and that she is perfectly able to drink the spoiled milk. Although there are many more ways

for things to be unequal, we don't need to canvas them. Millie's case can be filled out so that things are not unequal in any of these additional ways. That's because her case, as already described, looks like a paradigm case where other things are equal. Everything is running smoothly: Millie isn't confused about her beliefs or desires, she's capable of drinking the spoiled milk, and the world is cooperating.

Using a version of the Satisfaction–Disposition Principle with an 'other-things-equal' clause doesn't make a difference to our argument, since it makes perfect sense to think that other things are equal with Millie.

Zoom out for the moment and consider the broader dialectic. We have claimed that a certain case is possible, one where both P2 and P3 are true—where Millie wants to drink milk and is not disposed to drink the spoiled milk. Now we've added the supposition that other things are equal with Millie. But recall that for our argument to go through, we only need that there is *a* case where P2 and P3 are true and other things are equal. Our imagined interlocutor, on the other hand, must show that such a case (and all relevantly similar cases) is *impossible.*

4.7 In defense of the Satisfaction–Disposition Principle: on agent satisfaction vs. desire satisfaction

Another kind of worry about the Satisfaction–Disposition Principle doesn't concern the details of Millie's case, but rather the Satisfaction–Disposition Principle itself. You could grant the possibility of Millie's case as we've described it (that is, you could grant that it is possible that Millie wants to drink milk and is not disposed to drink the sour milk), yet deny that this shows anything about the satisfaction conditions of her desires. In this section we'll consider one objection to the Satisfaction–Disposition Principle; in the next section, another.

In arguing that desire satisfaction is *not* ways-specific (although they don't put it in those terms), Braun and Prinz distinguish desire satisfaction from what they call *agent satisfaction.* Desire satisfaction is a matter of whether some one or other of an agent's individual desires is satisfied; agent satisfaction is a matter of whether the agent herself *feels* satisfied.[18]

With this distinction in mind, you might worry that the thesis that agents are disposed to satisfy their desires has been misunderstood: the thesis should *not* be understood in terms of individual desire satisfaction (as it has been standardly thought (see e.g. §4.8 and the quotes on page 59)), but rather in terms of agent satisfaction. So the Satisfaction–Disposition Principle gets it wrong when it says that if you have *a* desire—an individual desire—that is satisfied in exactly the worlds where p is true, then you are disposed to what

[18]Unlike Prinz, who identifies agent satisfaction with an agent feeling satisfied, Braun does not explicitly say what he means by 'agent satisfaction'. We read him as having the same thing in mind as Prinz. Fara (2003), Persson (2005, ch. 10), and Lycan (2012) also discuss something like this distinction.

you believe will bring it about that p obtains. Rather, you are disposed to do what you believe will make yourself feel satisfied.

The worry is misguided. No doubt agents are in certain cases disposed to do what they believe will make themselves feel satisfied (although that doesn't mean they're not also disposed to do what they believe will satisfy their desires). But sometimes agents have desire-based dispositions that are *not* dispositions to do what they believe will make themselves feel satisfied. In such cases it's clear that desire satisfaction, not agent satisfaction, is what's at play.

Consider such a case: suppose that you want your name to live on after you die, and you do what you can to make it so. Suppose further that you don't in general feel good about merely *attempting* to reach your ends; rather, you feel satisfied only when you believe that your ends have been reached. (You're not one to hand out participation trophies.) As you work to make your name live on after your die—as you attempt to reach your end—you are unsure of whether you will succeed, and so you do not feel satisfied. And neither would you feel satisfied if you made your name live on after you die—if you in fact reached your end—since you don't feel anything at all after you die. You know all of this. So, as you do what you can to make your name live on, you neither experience nor anticipate any feeling of satisfaction.

You are disposed to do what you believe will make your name live on after you die. But your disposition is *not* to do what you believe will make yourself feel satisfied, since, again, you neither experience nor anticipate any feeling of satisfaction. Rather, your disposition is to do what you believe will satisfy one of your individual desires. The Satisfaction–Disposition Principle gets it right.

4.8 In defense of the Satisfaction–Disposition Principle: why accept it in the first place?

The final worry we'll consider about the Satisfaction–Disposition Principle is more general: why accept the Satisfaction–Disposition Principle in the first place?

The flat-footed answer is simple: the thesis that agents are disposed to satisfy their desires is true, and the Satisfaction–Disposition is a way of making this thesis precise. The subtler answer tells us why the Satisfaction–Disposition principle is a good way of making the thesis precise.

Recall how others have stated the thesis:

> [T]he primitive sign of having a desire is trying to satisfy it. (Humberstone (1990, p. 107), riffing on Anscombe)

> [T]he actions a desire is a disposition to perform are those that would satisfy that desire provided the agent's operative beliefs were true. (Stampe, 1994, p. 246)

> [A] desire is manifested in... behaviour aimed at satisfying the desire.

(Hyman, 2014, p. 85)

We can tease out two claims that are common among these quotes. The first is that from each desire, we can infer a disposition (or a trying, in Humberstone's case). The second is that this disposition is connected to the agent's desire in a certain way—it is a disposition to satisfy the desire. The Satisfaction–Disposition Principle, restated below, exemplifies both claims. It also allows us to make concrete predictions in a given case about whether an agent is disposed to do a certain thing, given her desires—something the above formulations don't allow us to do.

> *Satisfaction–Disposition Principle*
> If A has a desire that is satisfied in exactly the worlds where p is true, then A is disposed to do what she believes will bring it about that p obtains.

The crucial thing to establish is why this principle, and not some nearby principle, gets the connection between desires and dispositions right. Why would it be that it is exactly—i.e. all and only—the worlds where the desire is satisfied that matter to the disposition to satisfy it? Imagine that the principle were different.

Imagine, for example, that the principle were this: if A has a desire that is satisfied in *only* (but not necessarily all) worlds where p is true, then A is disposed to do what she believes will bring it about that p obtains. Then we would have a problem of disjunction introduction. Suppose Millie has a desire that is satisfied in exactly the worlds where she drinks fresh milk. She thereby has a desire that is satisfied only in worlds where she drinks fresh milk *or sprains her ankle*. She is not, though, disposed to do what she believes will bring it about that she drinks fresh milk or sprains her ankle.

Alternatively, imagine that the principle were this: if A has a desire that is satisfied in *all* (but not necessarily only) worlds where p is true, then A is disposed to do what she believes will bring it about that p obtains. Then we would have a problem of conjunction introduction. Suppose that Millie has a desire that is satisfied in exactly the worlds where she drinks fresh milk. She thereby has a desire that is satisfied in all worlds where she drinks fresh milk *and poisons her mother*. But Millie is not disposed to do what she believes will bring it about that she drinks fresh milk and poisons her mother.

The Satisfaction–Disposition Principle avoids both of these problems. Does it follow from the principle that Millie is disposed to do what she believes will bring it about that she drinks spoiled milk or sprains her ankle? No, because she does not have a desire that is satisfied in exactly the worlds where she does. Does it follow from the principle that Millie is disposed to do what she believes will bring it about that she drinks spoiled milk and poisons her mother? No, because she does not have a desire that is satisfied in exactly the worlds where she does.

4.9 Upshots: the dispositional role of desire satisfaction, revisited

We now have the premises, and so the conclusion: desire satisfaction is ways-specific. An agent may want p without having a desire that is satisfied in exactly the worlds where p is true.

This is a welcome conclusion: the thesis that desire satisfaction is ways-specific *explains* why agents are disposed to act as they are. Millie is not disposed to drink the spoiled milk *because* she is disposed to satisfy her desires and *she does not have a desire that is satisfied when she drinks the spoiled milk.* She has a desire-based disposition that is specific to certain ways of its being the case that she drinks milk because she has a desire whose satisfaction conditions are specific to certain ways of its being the case that she drinks milk. More generally, agents have ways-specific desire-based dispositions *because* they are disposed to satisfy their desires and *desire satisfaction is ways-specific.* (This prompts a question for the defender of the Satisfaction-is-Truth Principle: if desire satisfaction were *not* ways-specific, why would our desire-based dispositions be ways-specific, given that we're disposed to satisfy our desires?)

In addition to leading us to the conclusion that desire satisfaction is ways-specific, our argument gives us a new perspective on the dispositional role of desire satisfaction.

Consider, for example, that the following canonical principle connecting wanting and dispositions is false:

> *Want–Disposition Principle*
> If A wants p, then A is disposed to do what she believes will bring it about that p obtains.[19]

Millie wants to drink milk, but she is not disposed to drink the spoiled milk—not disposed to do what she believes will bring it about that she drinks milk. Millie has a ways-specific desire-based disposition, which the Want–Disposition–Principle says is impossible. Recall that if an agent has a ways-specific desire-based disposition, then for some p, (i) she wants p; (ii) there are ways for p to obtain that she is disposed to bring about; but (iii) there are other ways for p to obtain that she is not disposed to bring about, *even if she believes that she only can bring it about that p obtains in those ways.* If an agent has a ways-specific desire-based disposition, then the antecedent of the Want–Disposition Principle may be true of her, but the consequent not.

The Want–Disposition Principle is false, but in it is a kernel of truth. To see the kernel, consider that the Want–Disposition Principle is entailed by the conjunction of the Satisfaction-is-Truth Principle and the Satisfaction–Disposition Principle, repeated here.

[19] Audi (1973, p. 4), Davidson (1976, p. 243), and Stalnaker (1984, p. 15), among many others, advocate principles in this spirit.

Satisfaction-is-Truth Principle
If A wants p, then A has a desire that is satisfied in exactly the worlds where p is true.

Satisfaction–Disposition Principle
If A has a desire that is satisfied in exactly the worlds where p is true, then A is disposed to do what she believes will bring it about that p obtains.

Think of the Want–Disposition Principle as factored into these two principles that entail it. Once we remove the false part, the Satisfaction-is-Truth Principle, we are left with the kernel of truth, the Satisfaction–Disposition Principle. Agents are disposed to satisfy their desires.

Another flaw in the Want–Disposition Principle sheds further light on the dispositional role of desire satisfaction. If the Want–Disposition Principle were true (and remember, we don't think that it is), we should be able to determine, just on the basis of certain of an agent's beliefs and whether she wants p, whether she is disposed to bring it about that p obtains in some certain way. But we can't do this. If all we know about Millie is that she wants to drink milk and that she believes that the only milk that's available to her is the spoiled milk, we can't determine whether she's disposed to drink the spoiled milk. What we need to know is whether drinking the spoiled milk is a way for her desire to be satisfied. Only then will we be able to pin down Millie's disposition.

4.10 Upshots: wanting, desires, and the Fara–Braun debate

Readers familiar with the debate between Fara and Braun may wonder how our argument relates to the locus of that debate: a set of three principles on which Fara and Braun disagree. The first principle is a version of the influential Relational Analysis of attitude ascriptions (e.g. Stalnaker (1988), Schiffer (2003)) as applied to desire ascriptions. The second two concern wanting, desires, and how they're related to each other.[20]

First, some terminology. We assume that at the level of logical form, the complement of 'want' denotes a proposition, a standard assumption among semanticists (see e.g. Heim (1992) and von Fintel (1999)).[21] Let 'p' range over terms that denote propositions; let 'p' range over the corresponding propositions (ignoring any context-dependence in p); let 'A' range over the names of agents; and let 'A' range over the corresponding agents.

[20]There is a further question about what the noun 'desire' denotes—i.e. what desires are (as opposed to *wanting* or *desiring*). This question, discussed by e.g. Schroeder (2004) and Braun (2015), is, we believe, beyond the scope of our paper.

[21]This assumption is compatible with the thought that at the level of *surface form*, the complement of 'want' may not seem to denote a proposition—contrast e.g. 'Millie wants to drink milk' with 'Millie believes that she will drink milk'.

In stating the principles ourselves, we diverge slightly from Fara (2013)—she states all three principles as biconditionals, but her objection just concerns the left-to-right directions,[22] which is how we state them (and why we call them weak).

> *Weak Relational Analysis*
> If ⌜A wants p⌝ is true, then A stands in the relation denoted by 'wants' to p.[23,24]

> *Weak Content Component*
> If A stands in the relation denoted by 'wants' to p, then A has a desire with p as its content.[25]

> *Weak Specification Component*
> If A has a desire with p as its content, then A has a desire that is satisfied in exactly the worlds where p is true.

Fara rejects the conjunction of the principles; Braun accepts it.[26]

How do the three principles relate to what we've said? Their conjunction, plus the following overwhelmingly plausible quotation principle *entail the Satisfaction-is-Truth Principle.*

> *Quotation*
> If A wants p, then ⌜A wants p⌝ is true.[27]

We repeat the Satisfaction-is-Truth Principle again for reference:

> *Satisfaction-is-Truth Principle*
> If A wants p, then A has a desire that is satisfied in exactly the worlds where p is true.

We accept Quotation and thus side with Fara in rejecting the conjunction of the three principles.

Though we reject the conjunction of these principles, our argument is silent on which principle or principles should be rejected (our argument is compatible

[22]Braun makes the same point about the one of the principles, the Weak Specification Component, which we state just below.

[23]Stated more precisely, the principle is as follows. For all A, A, p, and p: if A denotes A and p denotes p, then if ⌜A wants p⌝ is true, then A stands in the relation denoted by 'wants' to p.

[24]Fara (2013) gives an instance of the principle: ' "Lora wants Rudy to be in London" is true just in case Lora bears the relation expressed by "wants" to the proposition that Rudy is in London' (p. 250). Braun states the principle as follows: 'If N is a proper name and S an infinitival phrase (with or without explicit subject), then ⌜N wants S⌝ is true iff the referent of N bears the relation expressed by "wants" to the proposition that S semantically expresses' (p. 144).

[25]For this principle and the next, see Fara's (2013) p. 253.

[26]More accurately, Braun accepts the latter two principles in conjunction with a different statement of the Weak Relational Analysis (see footnote 24).

[27]Stated more precisely, the principle is as follows. For all A, A, p, and p: if A denotes A and p denotes p, then if A wants p, then ⌜A wants p⌝ is true.

with rejecting any given one or combination of them). Determining which should be rejected requires settling broader questions in the philosophy of language and philosophy of mind, questions beyond the scope of this paper. We will, however, suggest a way to proceed.

Each principle links a certain fact about wanting, desires, or desire ascriptions to another. The Weak Relational Analysis, for example, links the proposition denoted by the complement of 'want' with a proposition to which the agent stands in the relation denoted by 'wants'. In particular, it says that the proposition denoted by the complement of a 'wants' ascription *is* a proposition to which the agent stands in the relation denoted by 'wants'. The Weak Content Component similarly says that the proposition to which the agent stands in the relation denoted by 'wants' *is* a proposition which is the content of one of the agent's desires. In turn, the Weak Specification Component says that the truth conditions of the proposition that is the content of the agent's desire *are* the satisfaction conditions of the agent's desires. All of the principles link various facts about wanting, desires, and desire ascriptions by saying that the propositions that figure in these facts are identical.

Our argument shows, though, that not all of these propositions can be identical. 'Millie wants to drink milk' is true, but Millie does not have a desire that is satisfied in exactly the worlds where she drinks milk. 'Millie wants to drink milk' is true but the truth conditions of the proposition denoted by the complement of 'want'—the proposition that Millie drinks milk—*are not identical to* the satisfaction conditions of any of Millie's desires. Rather, the relevant one of Millie's desires has satisfaction conditions that are *more specific* than this. That is to say, the satisfaction conditions of that desire are identical to the truth conditions of some proposition—perhaps the proposition that Millie drinks fresh milk—that *entails* the proposition that Millie drinks milk. Millie does not have a desire that is satisfied in exactly the worlds where she drinks milk, but she does (say) have a desire that is satisfied in exactly the worlds where she drinks fresh milk. Millie has a desire whose satisfaction conditions are ways-specific.

What we know, then, is that in attempting to link wanting, desires and desire ascriptions, at least one of the principles *underspecifies*—to use Fara's term—at least one of the relevant propositions. For example, it could be the Weak Content Component that goes wrong in this way. Then the proposition that is the content of the agent's relevant desire is more specific than the relevant proposition to which the agent stands in the relation denoted by 'wants'. If this is the case, we would propose replacing the Weak Content Component with the following principle: if A stands in the relation denoted by 'wants' to p, then, *for some proposition q that entails p*, A has a desire with q as its content.[28] Here, the proposition that is the content of the relevant

[28]Fara (2003, p. 159) advocates a similar principle: 'A desire (or related attitude) ascription of the form "A wants C" is true just in case A has a desire (or hope, etc.) with proposition Q as its exact content for some Q that entails the proposition expressed by the embedded clause C.' (For a related view, see what Condoravdi and Lauer (2016, p. 31)

one of the agent's desires is not identical to the relevant proposition (p) to which she stands in the relation denoted by 'wants'. Rather, it is a more specific proposition (q). It needn't be, of course, that the problem is with the Weak Component Component. One of the other two principles could be the culprit instead. In that case, we would propose to replace those principles with alternatives that capture the specificity of the relevant propositions.

4.11 Conclusion

Our argument has been this: agents are disposed to satisfy their desires; desire-based dispositions are ways-specific; so, desire satisfaction is ways-specific. The Satisfaction-is-Truth Principle, which entails that desire satisfaction is *not* ways-specific, is false. In reaching this conclusion, we sidestep concerns about the probative value of intuitions about when people get what they want—intuitions on which Fara and Lycan rely—appealing instead to principles concerning the relation between desires and dispositions to act.

Our argument opens up certain questions. Satisfaction is not truth, so what is it? Desire satisfaction is ways-specific, but to which ways? We must reject one of the three principles at issue in the debate between Fara and Braun, but which? Finally, is the satisfaction of other attitudes—hoping, dreaming, fearing—also ways-specific? We've given a template for how to answer: look first to the attitude's dispositional role, and then work your way back to satisfaction.

Whatever the answers to these questions are, our argument shows that there's an important disanalogy between desire and belief. The Truth-is-Truth Principle is true but the Satisfaction-is-Truth Principle is false. Desire satisfaction is not truth by another name.

call the 'Quine-Hintikka' analysis of 'want' ascriptions.) We believe that this is on the right track, but it's incorrect as it stands. It wrongly predicts that if ⌜A wants q⌝ is true, and q entails p, then ⌜A wants p⌝ is true. For example, it wrongly predicts that 'I want to die quickly' entails 'I want to die' (the example is from Anand and Hacquard (2013, p. 19)).

Bibliography

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

Arpaly, N. and Schroeder, T. (2013). *In Praise of Desire*. Oxford University Press.

Ashwell, L. (2017). Conflicts of desire: Dispositions and the metaphysics of mind. In Jacobs, J., editor, *Causal Powers*. Oxford University Press.

Audi, R. (1973). The concept of wanting. *Philosophical Studies*, 24(1):1–21.

Bach, K. (1994). Conversational impliciture. *Mind and Language*, 9(2):124–162.

Bach, K. (1997). Do belief reports report beliefs? *Pacific Philosophical Quarterly*, 78(3):215–241.

Bach, K. (2000). Quantification, qualification and context a reply to stanley and szabó. *Mind and Language*, 15(2&3):262–283.

Bach, K. (2001). You don't say. *Synthese*, 128(1):15–44.

Bach, K. (2005). Context *Ex Machina*. In Szabó, Z. G., editor, *Semantics Versus Pragmatics*, pages 15–44. Oxford University Press.

Baker, D. (2010). Ambivalent desires and the problem with reduction. *Philosophical Studies*, 150(1):37–47.

Blumberg, K. (2018). Counterfactual attitudes and the relational analysis. *Mind*, 127(506):521–546.

Blumberg, K. (ms). Beliefs, desires and descriptions. Unpublished manuscript.

Blumberg, K. and Holguín (ms). Embedded attitudes. Unpublished manuscript.

Bradley, R. (1999). Conditional desirability. *Theory and Decision*, 47(1):23–55.

Braun, D. (2011). Implicating questions. *Mind and Language*, 26(5):574–595.

Braun, D. (2015). Desiring, desires, and desire ascriptions. *Philosophical Studies*, 172(1):141–162.

Brown, J. (2004). *Anti-Individualism and Knowledge*. MIT Press.

Buchak, L. (2013). *Risk and Rationality*. Oxford University Press.

Buchak, L. (2014). Belief, credence, and norms. *Philosophical Studies*, 169(2):1–27.

Chandler, J. (2017). Descriptive decision theory. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy (Winter 2017 Edition)*.

Condoravdi, C. (2002). Temporal interpretation of modals - modals for the present and for the past. In *The Construction of Meaning*, pages 59–88. CSLI Publications.

Condoravdi, C. and Lauer, S. (2016). Anankastic conditionals are just conditionals. *Semantics and Pragmatics*, 9(8):1–69.

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

Dandelet, S. (ms). Partial desires, blinkered beliefs. Unpublished manuscript.

Daveney, T. F. (1961). Wanting. *Philosophical Quarterly*, 11(April):135–144.

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

Davidson, D. (1976). Hempel on explaining action. *Erkenntnis*, 10(3):239–253.

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

Dreier, J. (1996). Rational preference: Decision theory as a theory of practical rationality. *Theory and Decision*, 40(3):249–276.

Dreier, J. (2005). Pettit on preference for prospects and properties – discussion. *Philosophical Studies*, 124(2):199–219.

Dretske, F. (1988). *Explaining Behavior: Reasons in a World of Causes*. MIT Press.

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

Edgington, D. (1995). On conditionals. *Mind*, 104(414):235–329.

Erickson, P., Klein, J., Daston, L., Lemov, R., Sturm, T., and Gordin, M. (2013). *How Reason Almost Lost Its Mind: The Strange Career of Cold War Rationality*. University of Chicago Press.

Fara, D. G. (2003). Desires, scope, and tense. *Philosophical Perspectives*, 17(1):141–163. Originally published under the name 'Delia Graff'.

Fara, D. G. (2013). Specifying desires. *Noûs*, 47(2):250–272.

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

von Fintel, K. (ms). Conditional desires. http://kvf.me/cd. Unpublished slides.

von Fintel, K. and Iatridou, S. (2005). What to do if you want to go to Harlem: Anankastic conditionals and related matters. Ms., Massachusetts Institute of Technology.

von Fintel, K. and Iatridou, S. (2006). Anankastic conditionals and related matters. Slides of talk presented at the University of Tübingen.

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

Ganson, D. (2008). Evidentialism and pragmatic constraints on outright belief. *Philosophical Studies*, 139(3):441–458.

Grano, T. (2017). The logic of intention reports. *Journal of Semantics*, 43(4):587–632.

Grano, T. and Phillips-Brown, M. (ms). How belief-sensitive are *want* sentences? Unpublished manuscript.

Hawthorne, J. (2009). The lockean thesis and the logic of belief. In Huber, F. and Schmidt-Petri, C., editors, *Degrees of Belief*, pages 49–74. Synthese Library: Springer.

Heathwood, C. (2006). Desire satisfactionism and hedonism. *Philosophical Studies*, 128(3):539–563.

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

Huitink, J. (2008). *Modals, Conditionals and Compositionality*. PhD thesis, Nijmegen: Radboud University.

Humberstone, I. L. (1981). From worlds to possibilities. *Journal of Philosophical Logic*, 10(3):313–339.

Humberstone, I. L. (1990). Wanting, getting, having. *Philosophical Papers*, 99(August):99–118.

Hyman, J. (2014). Desires, dispositions and deviant causal chains. *Philosophy*, 89(1):83–112.

Jackson, F. (1985). Internal conflicts in desires and morals. *American Philosophical Quarterly*, 22(2):105–114.

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

Kratzer, A. (1981). The notional category of modality. In Eikmeyer, H.-J. and Reiser, H., editors, *Words, Worlds, and Contexts*, pages 38–74. de Gruyter.

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

Lasersohn, P. (1999). Pragmatic halos. *Language*, 75(3):522–551.

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.

Lewis, D. (1986). *On the Plurality of Worlds*. Blackwell.

Lycan, W. G. (2012). Desire considered as a propositional attitude. *Philosophical Perspectives*, 26(1):201–215.

Lycan, W. G. (ms). In what sense is desire a propositional attitude? Unpublished manuscript.

McDaniel, K. and Bradley, B. (2008). Desires. *Mind*, 117(466):267–302.

Moltmann, F. (2013). Propositions, attitudinal objects, and the distinction between actions and products. *Canadian Journal of Philosophy, Supplementary Volume on Propositions, Edited by G. Rattan and D. Hunter*, 43(5-6):679–701.

Montague, M. (2007). Against propositionalism. *Noûs*, 41(3):503–518.

Morillo, C. R. (1990). The reward event and motivation. *Journal of Philosophy*, 87(4):169–186.

Pearson, H. (ms). Attitude verbs. Unpublished manuscript for Semantics Companion.

Persson, I. (2005). *The Retreat of Reason: A Dilemma in the Philosophy of Life*. Oxford University Press.

Pettit, P. (1991). Decision theory and folk psychology. In Bacharach, M. and Hurley, S., editors, *Essays in the Foundations of Decision Theory*, pages 147–175. Blackwell.

Pettit, P. (2002). *Rules, Reasons, and Norms: Selected Essays.* Clarendon Press.

Phillips-Brown, M. (2018). I want to, but.... In Robert Truswell, Chris Cummins, C. H. B. R. and Rohde, H., editors, *Sinn und Bedeutung 21.*

Pollock, J. L. (2006). *Thinking About Acting: Logical Foundations for Rational Decision Making.* Oxford University Press USA.

Prinz, J. (ms). No satisfaction? the mundane truth about desires. Unpublished manuscript.

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 A. Arregui, M. L. R. and Salanova, A. P., editors, *Modality Across Syntactic Categories*, pages 610–633. Oxford University Press.

Sæbø, K. J. (1985). *Notwendige Bedingungen im Deutschen: zur Semantik modalisierter Sätze.* PhD thesis, University of Oslo.

Sæbø, K. J. (2001). Necessary conditions in a natural language. In Féry, C. and Sternefeld, W., editors, *Audiatur vox sapientiae: A Festschrift for Arnim von Stechow*, pages 427–449. Akademie-Verlagm.

Sæbø, K. J. (2017). Anankastic conditionals: If you want to go to harlem,.... Unpublished manuscript for the *Semantics Companion.*

Schiffer, S. R. (2003). *The Things We Mean.* Oxford University Press.

Schroeder, T. (2004). *Three Faces of Desire.* Oxford University Press.

Searle, J. R. (1983). *Intentionality: An Essay in the Philosophy of Mind.* Cambridge University Press.

Soames, S. (2005). Quantification, qualification and context: a reply to stanley and szabó. *Teorema: International Journal of Philosophy*, 24(3):7–30.

Soames, S. (2008). The gap between meaning and assertion: Why what we literally say often differs from what our words literally mean. In Soames, S., editor, *Philosophical Essays, Volume 1: Natural Language: What It Means and How We Use It*, pages 278–297. Princeton University Press.

Sperber, D. and Wilson, D. (1985). Loose talk. *Proceedings of the Aristotelian Society*, 86:153–171.

Staffel, J. (2016). Beliefs, buses and lotteries: Why rational belief can't be stably high credence. *Philosophical Studies*, 173(7):1721–1734.

Stalnaker, R. (1984). *Inquiry*. Cambridge University Press.

Stalnaker, R. (1988). Belief attribution and context. In Grimm, R. H. and Merrill, D. D., editors, *Contents of Thought*, pages 140–156. University of Arizona Press.

Stalnaker, R. (2008). *Our Knowledge of the Internal World*. Oxford University Press.

Stampe, D. W. (1986). Defining desire. In Marks, J., editor, *The Ways of Desire*. Blackwell.

Stampe, D. W. (1994). Desire. In Guttenplan, S. D., editor, *A Companion to the Philosophy of Mind*. Blackwell.

Stine, G. C. (1976). Skepticism, relevant alternatives, and deductive closure. *Philosophical Studies*, 29(4):249–261.

Strawson, G. (1994). *Mental Reality*. MIT Press.

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

Whyte, J. T. (1991). The normal rewards of success. *Analysis*, 51:65–73.

Williams, B. (1981). *Moral Luck: Philosophical Papers, 1973–1980*. Cambridge University Press.

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

Yablo, S. (2014). *Aboutness*. Princeton University Press.