(Counter)factual want ascriptions and conditional belief

Thomas Grano and Milo Phillips-Brown

Abstract. What are the truth conditions of want ascriptions? According to a highly influential and fruitful approach, championed by Heim (1992) and von Fintel (1999), the answer is intimately connected to the agent's beliefs: ⌜S wants p⌝ is true iff within *S*'s belief set, *S* prefers the *p* worlds to the ¬*p* worlds. This approach faces a well known and as-yet unsolved problem, however: it makes the entirely wrong predictions with what we call (counter)factual want ascriptions, wherein the agent either believes *p* or believes ¬*p*—e.g., 'I want it to rain tomorrow and that is exactly what is going to happen' or 'I want this weekend to last forever but of course it will end in a few hours'. We solve this problem. The truth conditions for want ascriptions are, we propose, connected to the agent's *conditional* beliefs. We bring out this connection by pursuing a striking parallel between (counter)factual and non-(counter)factual want ascriptions on the one hand and counterfactual and indicative conditionals on the other.

1 Introduction

What are the truth conditions of want ascriptions? According to a highly influential and fruitful approach, championed by Heim (1992) and von Fintel (1999), the answer is crucially intertwined with the agent's beliefs—specifically, with the set of worlds compatible with her beliefs, her belief set. The approach says, in outline, that ⌜S wants p⌝ is true just if *S* prefers the *p* worlds in her belief set to the ¬*p* worlds in her belief set.[1] (Theorists disagree about just what this preference relation amounts to.)

Originally developed to help account for presupposition projection (Heim, 1992; Geurts, 1998), the belief-set-based approach to want ascriptions has

[1]The symbol 'p' as a variable over syntactic objects that 'want' combines with is suggestive of 'proposition', and in this paper we will sometimes informally refer to the semantic values of those objects as 'propositions'. But nothing crucial hinges on this: all that matters for us is that objects of 'want' pick out sets of possible worlds. Insofar as prototypical want ascriptions in English have infinitival complements, we are thereby committed to the claim that infinitives pick out sets of possible worlds. We also treat objects of 'believe', which in English are prototypically finite clauses, as picking out sets of possible worlds. While there are some interesting semantic differences between infinitives and finite clauses, and accordingly, between objects of want ascriptions and objects of belief ascriptions, both contribute to the truth conditions of the sentences they inhabit in such a way that they can be fruitfully analyzed as picking out sets of possible worlds.

great theoretical power, and scholars continue to develop ever more sophisticated versions of it. As reviewed recently by Phillips-Brown (ms), the belief-set approach to want ascriptions has been shown to help explain the relationship between *want*, *wish*, and *be glad* (Heim 1992; see also section 9); the distinction between so-called 'predictive' and 'advisory' *want* (Jerzak, 2019); certain puzzles concerning *wish* (Blumberg, 2018) and *hope* (Blumberg ms);[2] various issues raised by Crnič (2011, Appendix A); and the interaction between *want* and conditionals (Jerzak, 2019; von Fintel, ms).[3]

Even given these successes, the belief-set approach faces a well-known problem: it yields the wrong results for ⌜S wants p⌝ whenever either (i) *S* believes *p* or (ii) *S* believes ¬*p*.[4] Call an ascription in case (i) a *factual want ascription*: the agent treats *p* as fact. In case (ii), we use the term *counterfactual want ascription*: the agent treats *p* as contrary to fact.

Example factual want ascription (FWA)[5]

(1) I want it to rain tomorrow (and I believe it will).

(adapted from Scheffler 2008)

Example counterfactual want ascription (CFWA)

(2) Wu wants to be promoted (but believes he won't be).

(adapted from Portner and Rubinstein 2012)

Let the umbrella term for FWAs and CFWAs be '(counter)factual want ascription', or '(C)FWA'.

In section 2, we'll present the problem of (C)FWAs in greater detail; for now, we'll just give you a feel for it, using (2) as an example. According to the belief-set approach, (2) is true just if Wu prefers the worlds in his belief set

[2]Blumberg's puzzles, and his solutions, are concerned with 'wish' and 'hope', but can be simply modified to apply to 'want'.

[3]Jerzak uses a graded-notion of belief in his semantics, while we concern ourselves in this paper with the all-out notion of belief found in the belief set; many of the benefits of Jerzak's graded-belief semantics can be enjoyed by an all-out-belief semantics. Jerzak also notes that with certain uses of 'want'—what he calls the *advisory* 'want'—it is not the agent's beliefs, but rather those in the context of evaluation or utterance that should figure in the semantics. We set such uses of 'want' to the side.

[4]Decision-theoretic approaches to want ascriptions (see e.g. Levinson 2003; Lassiter 2011; Phillips-Brown fc) face more or less this same problem. Wrenn (2010) proposes a solution that bears a certain resemblance to our own, although he confines his view to cases of all-things-considered desire (as opposed to 'some-things-considered' desire, in the sense of Phillips-Brown ms). Our approach, by contrast, is designed to be compatible with some-things-considered desire, either by bringing in multiple preference rankings (*à la* Levinson 2003; Crnič 2011), or by replacing possible worlds with something coarser (see section 7).

[5]See also Iatridou (2000:243) for several other examples of FWAs.

where he is promoted to… We needn't continue any further with the belief-set-based account, since it's broken down already: there are no worlds in Wu's belief set where he's promoted. He believes that he won't be promoted!

Theorists have noticed this problem before. Indeed, Heim herself saw it—and gave us the following, memorable CFWA, (3), below—when she brought the belief-set-based approach onto the scene.

(3) I want this weekend to last forever (but of course I know it will be over in a few hours). (Heim 1992, p. 199)

Yet Heim did not venture a solution.[6] To our knowledge, only one theorist has: Rubinstein (2017). But her solution is not satisfactory, as we argue in section 3.

In sections 4–5, we develop a solution of our own—one in which *conditional belief* is integral to the semantics of 'want'. Our account is inspired by a striking parallel between indicative vs. counterfactual conditionals on the one hand and non-(C)FWAs vs. (C)FWAs on the other. By way of preview, our proposal, stated informally, is that ⌜S wants p⌝ is true iff *S* prefers what she believes will (would) happen if *p* is (were) true to what she believes will (would) happen if ¬*p* is (were) true. As we will show, this approach enjoys all the same virtues of the belief-set-based approach for non-(C)FWAs while endowing the semantics with just enough flexibility to handle (C)FWAs in a tightly constrained way. After presenting our solution, we turn in sections 6–9 to some further contextualizing considerations and extensions before concluding the paper.

2 The problem, formalized

We'll bring the problem of (C)FWAs into clearer view with an instance of the belief-set-based approach to want ascriptions: von Fintel's (1999) best-worlds semantics, which is patterned after Kratzer's (1981, 1991) classic system of modals.

We said that on the belief-set-based view, ⌜S wants p⌝ is true iff *S* prefers the *p* worlds in her belief set to the ¬*p* worlds in her belief set. Von Fintel's se-

[6] Heim (1992) also noticed the problem of FWAs, with the sentence '(John hired a babysitter because) he wants to go to the movies tonight' (p. 199). She proposed to solve this problem by replacing the belief set with the set of worlds "compatible with everything that [the agent] in [the evaluation world] believes to be the case no matter how he chooses to act." Insofar as John's belief that he will go to the movies tonight is underpinned by his intention to do so, this particular example is no longer a problem. But this solution does not extend to all factual want ascriptions, like (1) above.

mantics exemplifies this view by saying that ⌜S wants p⌝ is true iff p is true in all of the best worlds in S's belief set, as ranked by S's preferences.

More formally, let Bel_S be S's belief set, and let best_S be a function that takes a set of worlds Γ and returns the subset of Γ that is best according to S's preferences:[7]

(4) *von Fintel's semantics*[8]
$[\![\text{S wants p}]\!] = 1$ iff $\forall w \in \text{best}_S(\text{Bel}_S)\colon p(w) = 1$

Absent any caveats, von Fintel's semantics would predict that FWAs are vacuously true: because the agent believes p, there are only p worlds in Bel_S; the best worlds in Bel_S are thus p worlds. Conversely, CFWAs would be vacuously false: because the agent believes $\neg p$, p worlds are absent from Bel_S and thereby absent from Bel_S's best worlds.[9]

In response to this threat of vacuous truth and vacuous falsity, proponents of the belief-set approach, von Fintel included, have hypothesized that (C)FWAs suffer from presupposition failure: $[\![\text{S wants p}]\!]$ is undefined whenever S either believes p or believes $\neg p$.[10] But we already know this to be wrong: the (C)FWAs in section 1—i.e. (1), 'I want it to rain tomorrow (and I believe it will)' and (2), 'Wu wants to be promoted (but believes he won't be)'—are intuitively felicitous and contingent want ascriptions.[11]

With or without a presuppositional component, then, von Fintel's semantics goes wrong. Where to go from here? To map a way forward, we propose to see von Fintel's semantics as breaking into three assumptions that jointly entail it. Begin with a basic assumption of the best-worlds semantics, one that's shared by all of the other belief-set-based semantics, and one that we will pre-

[7] Let $w' <_S w$ mean that w' is better than w according to S's preferences: $\text{best}_S(\Gamma) = \{w \in \Gamma \mid \neg\exists w' \in \Gamma\colon w' <_S w\}$.

[8] We make an innocuous simplification here: following Heim (1992), von Fintel (1999) actually uses not Bel_S as his domain but rather a certain superset of it, Bel^*_S, which screens off S's beliefs grounded in her own intentions: $\text{Bel}^*_S = \{w \mid w$ is compatible with what S believes will happen no matter how she chooses to act$\}$. As discussed in footnote 6, this is in order to account for a certain species of FWAs in which the agent intends to make it the case that the desired proposition holds. But since not all FWAs are of this species, the problem of FWAs persists, as of course does the problem of CFWAs.

[9] We're assuming that Bel_S, and therefore $\text{best}_S(\text{Bel}_S)$, is non-empty. If Bel_S is instead empty, then counterfactual want ascriptions are vacuously true: all of the best worlds in Bel_S (there are none!) are vacuously p worlds.

[10] The details of this presuppositional component differ by theorist. See e.g. Heim 1992: 198 or von Fintel 1999:117.

[11] This is not to say that *all* (C)FWAs are impeccable, and indeed we speculate that competition from the explicitly counterfactual 'wish' and the explicitly factual 'be glad (that)' may render (C)FWAs weakly dispreferred under some conditions. See section 9.

serve. The assumption is that 'want' ascriptions are evaluated against some domain of worlds, D, and the agent's preferences are among the worlds in D.

(5) *Form of a semantics for 'want'*
$[\![\text{S wants p}]\!] = 1$ iff S prefers the p worlds in D to the $\neg p$ worlds in D.

This form requires us to fill in two elements: what it is for S to prefer p to $\neg p$ in D, and what D is. The belief-set-based view says that D is the belief set:

(6) *Belief-set domain*
$D = \text{Bel}_S$

And we can read von Fintel as saying that S prefers the p worlds in D to the $\neg p$ worlds in D just if all of the best worlds in D are p worlds, as ranked by S's preferences. Or, more formally:

(7) *Best-worlds preference within a domain*
S prefers the p worlds in D to $\neg p$ worlds in D iff $\forall w \in \text{best}_S(D)$:
$p(w) = 1$.

Together, (5), (6), and (7) entail von Fintel's semantics. Von Fintel's semantics must be rejected, and so one of these three views must be rejected. Certainly one could reject (5) and instead take a different approach to 'want'. Some have done so.[12] We would like to know if the (C)FWA problem can be solved—and so the theoretical fruits of the belief-set-based approach enjoyed—while holding on to (5), and indeed we believe that it can be, as we'll argue in section 5.

So: we must reject either either the Belief-set domain (i.e. (6)) or the Best-worlds preference within a domain (i.e. (7)). Not necessarily (7). While its adequacy is debated, rejecting it is not a solution to the problem of (C)FWAs. If we were to replace it with something else, like a Heimian notion of preference within a domain (see section 6), we'd be playing a game of whack-a-mole: the (C)FWA problem would pop up with its replacement. Out with (6), then—the domain for (C)FWAs is not the belief set. So what is it? In the next section we review and argue against Rubinstein's answer to this question; in sections 4–5 we develop our own.

[12] For example, the Content Specification version of the so-called Relational Analysis—an influential semantics in the philosophy literature (Fara, 2013)—quantifies over the agent's *desires* (rather than over some domain of worlds). Similarly, Condoravdi and Lauer (2016) quantifies over the agent's *preferences*.

3 Against Rubinstein's domain

Rubinstein (2017) proposes a variant of Heim's (1992) semantics for want ascriptions, amended so as to carry out a "relaxed reliance of belief in desire statements" (p. 110). It is important to note at the outset that Rubinstein's framing motivation is not the problem of (C)FWAs, but rather a separate problem for want ascriptions identified by Villalta (2008), which we have our own take on and will discuss in due course in section 7. However, as Rubinstein herself notes, and as we shall see presently, her proposal does have some relevance to (C)FWAs.

Rubinstein offers her proposal as a Heim 1992-style re-implementation of Villalta's (2008) context-dependent semantics for want ascriptions. In a nutshell, Rubinstein's proposal is that when we evaluate ⌜S wants p⌝, the domain is not *S*'s belief set, but rather a contextually circumscribed domain that is defined only when its intersection with *S*'s belief set is *diverse* with respect to *p* (in the sense of (Condoravdi, 2002)): that is, (i) its intersection with *S*'s belief set and with *p* is non-empty and (ii) its intersection with *S*'s belief set and with ¬*p* is non-empty (p. 117):

(8) *Rubinstein's (2017) Domain for* ⌜S wants p⌝
$D \cap \text{Bel}_S \cap p \neq \emptyset$ and $D \cap \text{Bel}_S \cap \neg p \neq \emptyset$
(where *D* is contextually circumscribed)

As (8) stands, it will not help with the problem of (C)FWAs, because although it relaxes the relationship between *D* and Bel_S by merely requiring them to have a non-empty overlap instead of equating them, it still requires non-empty overlap between Bel_S and *p* (dooming CFWAs) and between Bel_S and ¬*p* (dooming FWAs). But to give the general idea a fighting chance, it will be instructive to explore two variations on (8) that remove this requirement while leaving much of the rest of the proposal intact.

The first variation, suggested to us by [redacted for review] (pers. comm.), is to revise (8) in such a way that *D* must be diverse with respect to *p* and *D* must have a non-empty overlap with both *p* and ¬*p*, but there is no requirement imposed directly on the relationship between Bel_S and *p* or ¬*p*, as in (9). This allows for (C)FWAs.

(9) *Variation 1 on Rubinstein's (2017) Domain for* ⌜S wants p⌝
$D \cap \text{Bel}_S \neq \emptyset$
$D \cap p \neq \emptyset$ and $D \cap \neg p \neq \emptyset$
(where D is contextually circumscribed)

The second variation is to remove any reference whatsoever to Bel_S. This is suggested by Rubinstein herself: citing Heim's never-ending weekend sentence (see (3) above), Rubinstein says "one might…challenge the very assumption that beliefs semantically restrict desire statements at some level" (p. 119). That would give us (10), which requires merely that D be diverse with respect to p:

(10) *Variation 2 on Rubinstein's (2017) Domain for* ⌜S wants p⌝
$D \cap p \neq \emptyset$ and $D \cap \neg p \neq \emptyset$
(where D is contextually circumscribed)

In what follows, we will show that (9) goes wrong because the relationship between D and Bel_S is insufficiently constrained. Since (10) leaves that relationship even more unconstrained (in fact, completely unconstrained), it follows that (10) goes wrong too.

We proceed by first elaborating on what Rubinstein says in prose about the relationship between D and Bel_S. Rubinstein suggests that in general, the relation between D and *S*'s belief set "may be one of inclusion, such that the accessible worlds are a superset of the doxastic alternatives, arrived at by potentially suspending some of the subject's beliefs" (p. 118). This suggestion is quite similar to Heim's (1992) reaction to the problem of CFWAs; Heim says, "maybe for some reason not all the subject's beliefs are taken into account" (p. 200). We believe that there is something importantly right about this idea. Crucially, though, its ultimate success depends on whether we can answer the following question: what are the rules that govern which worlds D contains, beyond those in *S*'s belief set? Stated differently: what are the rules that tell us which of *S*'s beliefs to suspend in arriving at D? Is it really left entirely to context? The only guideline Rubinstein offers is in connection with one particular example she considers (an example from Heim 1992, 'I want to teach Tuesdays and Thursdays next semester'), where she says that the domain consists of "circumstantially accessible worlds in which I teach next semester… Only a proper subset of these are the belief worlds of the subject" (p. 118).[13] Can we then say

[13] Rubinstein makes a similar suggestion in earlier work (2012), where, in connection with Heim's never-ending weekend sentence, she says (emphasis added), "What such examples show is that the possibilities that are relevant for a desire statement may be possibilities that are *circumstantially accessible*, yet doxastically inaccessible" (p. 116).

that, in general, *D* is the union of *S*'s belief set with some contextually circumscribed subset of circumstantially accessible worlds?

This is not right for the problem of (C)FWAs (nor is it right in general). Imagine, in the promotion case, that Wu is eager for new responsibilities. He believes that he'd feel invigorated with the tasks required by the higher position. (2), 'Wu wants to be promoted', is true. But you can easily imagine a way of filling out Wu's case so that Rubinstein's semantics will erroneously predict that (2) is false. Imagine, for example, a context in which, unbeknownst to Wu, there are envious assassins, lying in wait, who will poison Wu and his family, should he be promoted. If (2) is to be predicted true, promotion-worlds must be best (most desirable to Wu) in *D* (that's how the Best-worlds Semantics works). For Rubinstein, *D* is Bel_{Wu}—which doesn't contain promotion-worlds, since Wu believes he won't be promoted—plus contextually circumscribed, circumstantially accessible worlds in which he is promoted. For Rubinstein, as with a best-worlds semantics, ⌜S wants p⌝ is true only if *p* is true in all of the best worlds in the domain (p. 117). But these circumstantially accessible worlds in the domain, those that exceed Bel_{Wu}, are not best; they're as bad as can be! In them, Wu and his family are poisoned. (2) is (wrongly) predicted false.

In more general terms, because an agent may be entirely wrong about the world he inhabits (as e.g. Wu is), contextually circumscribed, circumstantially accessible worlds (e.g. secret-assassin worlds) may be entirely untethered from the agent's beliefs (e.g. Wu may believe as strongly as he believes anything that there aren't any secret assassins). Rubinstein's mistake is allowing *D* to contain worlds so distant from the agent's beliefs, since such worlds are in general irrelevant to what an agent wants.

At this point one might try to salvage Rubinstein's approach by adjusting her proposal to be about worlds that are circumstantially accessible, not relative to the actual world, but rather relative to the agent's beliefs.[14] But this just duplicates the (C)FWA problem: worlds that one does not believe to be possible cannot be ones that ones believes are circumstantially accessible—after all, circumstantial accessibility is *realistic* in Kratzer's (1981) sense. What we need instead for the domain is something more along the lines of "worlds that the agent believes *would* be circumstantially accessible in certain doxastically counterfactual worlds consistent with the desired proposition." And this moves us in the direction we take below.

[14] In fact, Rubinstein (2017), in footnote 11 on p. 117, mentions the possibility of "[a]nchoring the modality to the subject," but does not pursue it.

4 Interlude: Modeling conditional belief

As we mentioned in the introduction, our solution to the problem of (C)FWAs makes key use of the agent's *conditional beliefs*. To get a handle on just what we have in mind, we will first propose a certain way of modeling conditional belief. To be sure, there are myriad ways to understand conditional belief other than the one we present here (see e.g. Edgington 2014 section 3.1 and references therein), and we have no aspirations to supplant them nor to advance a full-blown theory. We are merely exploring a certain way of seeing conditional belief that furthers our understanding of the meaning of 'want'.

Our jumping-off point is Stalnaker's (1968) system of conditionals, on which ⌜if p, q⌝ is true at w just if q is true at the p world that's most similar to w, or $\text{Sim}_w(p)$.[15] When ⌜if p, q⌝ is indicative, you can understand $\text{Sim}_w(p)$ as representing what will happen, relative to w, if p is true. Similarly, when ⌜if p, q⌝ is counterfactual, $\text{Sim}_w(p)$ represents what *would* happen, relative to w, if p *were* true. To cover both cases, say that $\text{Sim}_w(p)$ represents what will (would) happen if p is (were) true.

Understanding $\text{Sim}_w(p)$ in this way facilitates a parallel understanding of conditional belief. For each w in Bel_S, find the most similar p world to w: collect these p worlds into a set, which we'll call '$\text{ConBel}_S(p)$':[16]

(11) $\quad \text{ConBel}_S(p) =_{df} \{w' \mid \exists w \in \text{Bel}_S\colon \text{Sim}_w(p) = w'\}$

We invite you to understand $\text{ConBel}_S(p)$ as representing what S believes will (would) happen if p is (were) true.

We turn now to a further development of Stalnaker's theory of conditionals, one that will end up playing a crucial role in our solution to the problem of (C)FWAs. In particular, Stalnaker (1975) is concerned with the role of conditionals in conversation—with how conditionals interact with the *context* set of a conversation, the set of worlds compatible with the presuppositions of the conversation's participants. Here's Stalnaker:

[15] While the value of Sim is a single world, the work to which we put Sim to use can be done just as well by a Lewis 1973-style similarity function whose value is a set of worlds.

[16] We are grateful to [redacted for review] (pers. comm.) for pointing out to us that $\text{ConBel}_S(p)$ is essentially the same as what Lewis (1976) would call the "image" of Bel_S on p. [redacted for review] also notes that Gardenfors (1982) generalizes Lewis's notion of imaging in a way that might enable us to carry out our conditional belief account of want ascriptions without relying on Stalnaker's theory of conditionals, though for the sake of concision we do not explore that option here.

> [I]f the conditional is being evaluated at a world in the context set, then the world selected must, if possible, be within the context set as well [...] In other words, all worlds within the context set are closer [i.e. more similar] to each other than any worlds outside it. The idea is that when a speaker says 'If *A*', then everything he is presupposing to hold in the actual situation is presupposed to hold in the hypothetical situation in which *A* is true. [...] it is at least a normal expectation that the selection function should turn first to [the worlds in the context set] before considering *counterfactual* worlds—those presupposed to be non-actual. (Stalnaker 1975, pp. 275–6)

You can translate Stalnaker's thought into our formal vocabulary as follows, where C is the context set:

(12) *Stalnaker's Context Set Constraint*
$\forall w \in C, \forall p$: if C contains p worlds, then $\text{Sim}_w(p) \in C$

Stalnaker uses this constraint to characterize indicative vs. counterfactual conditionals. A conditional ⌜if p, q⌝ is not indicative or counterfactual full stop; rather, it is counterfactual or indicative *with respect to a context set*. ⌜if p, q⌝ is counterfactual with respect to C just if Sim_w must reach outside C to find a p world, which according to the constraint, happens only when C doesn't contain p worlds—that is, when the conversational participants treat p as contrary to fact. In contrast, ⌜if p, q⌝ is indicative with respect to C when the conversational participants don't treat p as contrary to fact.

The belief set, we suggest, is a natural analogue of the context set. The context is the set of worlds compatible with the presuppositions of the conversational participants—in other words, those worlds that the participants treat as live possibilities. Similarly, the belief set is the set of worlds compatible with the beliefs of the agent; in other words, those worlds that the agent treats as live possibilities.

This analogy between belief set and context set is the foundation for a constraint on the belief set that's analogous to Stalnaker's Context Set Constraint. Adapting the passage from Stalnaker:

> All worlds within the belief set are more similar to each other than any worlds outside it. The idea is that when a conditional 'If *A*', is entertained with respect to the agent's beliefs, everything the agent

> believes to hold in the actual situation holds in the hypothetical situation in which A is true. [...] it is at least a normal expectation that the selection function should turn first to the worlds in the belief set before considering *counterfactual* worlds—those believed by the agent to be non-actual.
>
> (us, riffing on Stalnaker)

Formally, we have:

(13) *Stalnakerian Belief Set Constraint*
$\forall w \in \text{Bel}_S$, $\forall p$: if Bel_S contains p worlds, then $\text{Sim}_w(p) \in \text{Bel}_S$

Just as Stalnaker used his constraint to characterize indicative and counterfactual conditionals, we will use the Stalnakerian constraint to characterize indicative and counterfactual conditional belief. As you know, we represent an agent's conditional beliefs about p with $\text{ConBel}_S(p)$, which abbreviates $\{w' \mid \exists w \in \text{Bel}_S\colon \text{Sim}_w(p) = w'\}$.

When S believes $\neg p$, $\text{ConBel}_S(p)$ represents *S's counterfactual conditional beliefs*: what S believes *would* happen if p *were* true. This is because when S believes $\neg p$—i.e. when Bel_S contains no p worlds—Sim_w must reach outside Bel_S to find p worlds. In other words, $\text{ConBel}_S(p)$ is constructed only of worlds that S believes are contrary to fact.

Conversely, when S's beliefs are consistent with p, $\text{ConBel}_S(p)$ represents *S's indicative conditional beliefs*: what S believes *will* happen if p *is* true. That's because when S's beliefs are consistent with p—i.e. when Bel_S contains p worlds—the Stalnakerian Belief Set Constraint dictates that $\text{ConBel}_S(p)$ doesn't extend beyond Bel_S. $\text{ConBel}_S(p)$ contains only worlds in Bel_S, which is to say it doesn't contain worlds contrary to what S believes is fact.

One note before moving on. We have, following an influential tradition, characterized the indicative–counterfactual distinction in terms of the context set. This characterization tends to be rejected, though, by those who understand counterfactuals in terms of the past tense (e.g. Arregui (2007), Ippolito (2013)). Adopting this alternative understanding may lead one to reject our *interpretation* (which uses counterfactuals) of the machinery we use below to construct the domain for 'want'. But we believe that adopting this alternative understanding does not thereby give one reason to reject the domain itself, or the predictions that one gets if one adopts it.

5 The Conditional-belief Domain: Solving the (C)FWA problem

Equipped with the approach to conditional belief from section 4, we can solve the problem of (C)FWAs. Recall the basic set-up from section 2. We have three views that jointly entail von Fintel's best-worlds semantics:

(5) *Form of a Semantics for 'Want'* (we accept)
$[\![\text{S wants p}]\!] = 1$ iff S prefers the p worlds in D to the $\neg p$ worlds in D.

(6) *Belief Set Domain* (we reject)
$D = \text{Bel}_S$

(7) *Best-worlds Preference in a Domain* (we're ecumenical)
S prefers the p worlds in D to the $\neg p$ worlds in D iff $\forall w \in \text{best}_S(D)$: $p(w) = 1$.

As we've discussed, the solution to the problem of (C)FWAs lies in replacing the Belief Set Domain. In proposing a replacement below, we presuppose (5). But we are ecumenical on just what preference within a domain is. Our domain is compatible with the Best-worlds conception (i.e. (7)) and, for example, with Heim's conception (see section 6).

Our domain is the union of $\text{ConBel}_S(p)$ and $\text{ConBel}_S(\neg p)$. In other words, the domain represents S's conditional beliefs about p and S's conditional beliefs about $\neg p$.

(14) *Conditional Belief Domain* (our proposal)
$D = \text{ConBel}_S(p) \cup \text{ConBel}_S(\neg p)$
$= \{w' \mid \exists w \in \text{Bel}_S\colon \text{Sim}_w(p) = w'\} \cup \{w' \mid \exists w \in \text{Bel}_S\colon \text{Sim}_w(\neg p) = w'\}$

This domain works for (C)FWAs and non-(C)FWAs alike. In particular, it makes the same predictions for non-(C)FWAs as the standard semantics for 'want', while correcting the mistakes of that semantics for (C)FWAs. Below, we'll examine non-(C)FWAs and (C)FWAs in turn.

5.1 Non-(C)FWAs

Our truth conditions for non-(C)FWAs turn on the interaction between the agent's preferences and her *indicative* conditional beliefs:

(15) *Truth conditions for non-(C)FWAs, informally*
If S believes neither p nor $\neg p$, ⌜S wants p⌝ is true iff:
S prefers what she believes *will* happen if p *is* true to what she believes *will* happen if p *is* false.

For example, suppose that Li is unsure if there are any copies of *War and Peace* available, and therefore unsure if she will be able to read it. Li neither believes that she'll read *War and Peace* nor believes that she won't. And so (16) just below is a non-(C)FWA.

(16) Li wants to read *War and Peace*.

On our account, (16) is true just if Li prefers what she believes will happen if she does (her friends will think she's an intellectual) to what she believes will happen if she doesn't (they'll think she's boring).

Here's how these informal truth conditions capture the formalism above. Because we're dealing with a non-(C)FWA, Li's beliefs are compatible with her reading *War and Peace* and compatible with her not doing so. Both $\text{ConBel}_{Li}(War)$ and $\text{ConBel}_{Li}(\neg War)$—i.e. our domain for (16)—are indicative with respect to Bel_{Li}. (Where *War* is the proposition that Li reads *War and Peace*.) This is to say that $\text{ConBel}_{Li}(War)$ and $\text{ConBel}_{Li}(\neg War)$ respectively represent what Li believes will happen if she reads *War and Peace* and what she believes will happen if she doesn't.

That's how the agent's beliefs enter the picture. Now consider her preferences. The form of a semantics for 'want' that we endorse, (5), says that ⟦S wants p⟧ = 1 iff *S* prefers the *p* worlds in *D* to the ¬*p* worlds in *D*. For us, *D* is the union of $\text{ConBel}_S(p)$ (a set of *p* worlds) and $\text{ConBel}_S(\neg p)$ (a set of ¬*p* worlds). The form of the semantics then becomes this: ⟦S wants p⟧ = 1 iff *S* prefers the worlds in $\text{ConBel}_S(p)$ to the worlds in $\text{ConBel}_S(\neg p)$, which is to say that *S* prefers the worlds that represent what she believes will happen if *p* is true to the worlds that represent what she believes will happen if *p* is false. This matches (15).

We've said that with non-(C)FWAs, our view makes the same predictions as the best-set-based view. That's because with non-(C)FWAs, our domain *is* the belief set.

To show this, we'll first establish a key fact:

(17) If Bel_S contains *p* worlds, then ConBel_S contains all and only the *p* worlds in Bel_S.

$\text{ConBel}_S(p)$ contains *all* the *p* worlds in Bel_S because we place a so-called *centering* constraint on the Sim function: for any proposition *p* and world *w*, if *w* is a *p* world, then $\text{Sim}_w(p) = w$. (See e.g. Stalnaker 1968; the constraint is motivated by the thought that a world is more similar to itself than to any other world.) And so if *w* is a *p* world in Bel_S, then *w* is in $\text{ConBel}_S(p)$.

$\text{ConBel}_S(p)$ contains *only* the p worlds in Bel_S because of the Stalnakerian Belief Set constraint. With non-(C)FWAs, there are p worlds in Bel_S. The Stalnakerian Belief Set Constraint then dictates that $\text{Sim}_w(p)$—and therefore $\text{ConBel}_S(p)$—never reaches outside of Bel_S, for all w in Bel_S. So if w is a p world in $\text{ConBel}_S(p)$, then w is in Bel_S.

With (17) established, return to our domain, $\text{ConBel}_S(p) \cup \text{ConBel}_S(\neg p)$. With a non-(C)FWA, there are p worlds in Bel_S, so $\text{ConBel}_S(p)$ contains all and only the p worlds in Bel_S; with a non-(C)FWA, there are $\neg p$ worlds in Bel_S, so $\text{ConBel}_S(\neg p)$ contains all and only the $\neg p$ worlds in Bel_S. Put these two facts together and $\text{ConBel}_S(p) \cup \text{ConBel}_S(\neg p)$ contains all and only the worlds in Bel_S; in other words, our domain for non-(C)FWAs is Bel_S.

5.2 (C)FWAs

While non-(C)FWAs operate entirely with the agent's indicative conditional beliefs, (C)FWAs operate half with the agent's indicative conditional beliefs and half with her counterfactual conditional beliefs.

Begin with CFWAs.

(18) *Truth conditions for CFWAs, informally*
If *S* believes $\neg p$, ⌜S wants p⌝ is true iff:
S prefers what she believes *would* happen if *p were* true to what she believes *will* happen given that $\neg p$ *is* true.

Take, for instance, our CFWA from before, (2), repeated below:

(2) Wu wants to be promoted (but he believes he won't be).

(2) is true just if he prefers what he believes would happen if he were to be promoted (this is the counterfactual element) to what he believes will happen given that he won't be promoted (this is the indicative element[17]).

These informal truth conditions capture the above formalism as follows. $\text{ConBel}_{Wu}(\textit{Promoted})$ is *counterfactual* with respect to Wu's beliefs because he believes that he won't be promoted: $\text{ConBel}_{Wu}(\textit{Promoted})$ represents what Wu believes would happen if he were to be promoted. And $\text{ConBel}_{Wu}(\neg \textit{Promoted})$ is *indicative* with respect to Wu's beliefs because his beliefs are compatible with the possibility that—in fact, entail that—he won't be promoted:

[17]We informally paraphrase the indicative element as "*given that* he won't be promoted" rather than "*if* he isn't promoted", because the latter tends to lead to the inference that the agent considers it to be possible that he will be promoted, which is not the case since this is a CFWA. Here and in what follows, we use "given that p" in situations where p is entailed by the agent's belief set.

$\text{ConBel}_{Wu}(\neg \textit{Promoted})$ represents what Wu believes will happen given that he won't be promoted.

Put another way, $\text{ConBel}_{Wu}(\textit{Promoted}) \cup \text{ConBel}_{Wu}(\neg \textit{Promoted})$ consists of (i) the worlds where Wu is promoted that are maximally similar to worlds in Bel_{Wu} and (ii) ConBel_{Wu} itself. (i) holds because there aren't any *Promoted* worlds in Bel_{Wu}, so the Sim function must reach outside of Bel_{Wu} to find *Promoted* worlds; (ii) holds because of (17) and the fact that Bel_{Wu} contains only $\neg$*Promoted*-worlds.

The domain for CFWAs, more generally, comprises (i) the p worlds maximally similar to worlds in Bel_S and (ii) Bel_S itself. By adopting this domain, and an analogous one for FWAs (see just below), our domain embodies—in a highly constrained way—Heim's (1992) and Rubinstein's (2017) idea that want ascriptions are sometimes evaluated against a superset of the agent's belief set.

Turn now to FWAs, whose truth conditions are the mirror image of those for CFWAs:

(19) *Truth conditions for CFWAs, informally*
If *S* believes p, ⌜S wants p⌝ is true iff:
S prefers what she believes *will* happen given that *p is* true to what she believes *would* happen if $\neg p$ *were* true.

Take our FWA from above, repeated here:

(1) I want it to rain tomorrow (and I believe it will).

On our view, (1) is true just if I prefer what I believe *will* happen given that it will rain tomorrow (the boring company picnic will be canceled) to what I believe *would* if it didn't rain tomorrow (I'd have to go to the picnic).

To see why these are the truth conditions, run the mirror of the reasoning we ran for CFWAs. What you'll conclude is this: While the domain in general for CFWAs is Bel_S plus the p worlds maximally similar to Bel_S (Bel_S is counterfactual with respect to *S*'s beliefs), the domain for FWAs is Bel_S plus the $\neg p$ worlds maximally similar to Bel_S (Bel_S is factual with respect to p).

5.3 Summing up

Let us zoom out. Our proposed domain is $\text{ConBel}_S(p) \cup \text{ConBel}_S(\neg p)$. The makeup of this set differs among non-(C)FWAs, CFWAs, and FWAs. Specifically, the domain takes the following forms, as we saw in sections 5.1 and 5.2:

ascription type	our domain
non-(C)FWA	Bel_S
FWA	Bel_S and $\neg p$ worlds maximally similar to Bel_S
(C)FWA	Bel_S and p worlds maximally similar to Bel_S

The differences between these domains lie in the differences in whether p and $\neg p$ are respectively indicative or counterfactual with respect to the agent's beliefs. In all three of the cases, the agent's conditional beliefs form the core of the truth conditions, which we summarize in (20). (Recall that for any proposition q, $\text{ConBel}_S(q)$ represents what S believes will (would) happen if q is (were) true.)

(20) *Truth conditions for 'want', informally*
⌜S wants p⌝ is true iff:
S prefers what she believes will (would) happen if p is (were) true to what she believes will (would) happen if $\neg p$ is (were) true.

While this fulfills our core goal of solving the problem of (C)FWAs, we would like, before closing, to address a few issues that will help to better contextualize our proposal with respect to other ideas, problems, and facts found in the literature. In particular, we will: clarify the relationship between the role that conditionals play in our semantics vs. the role they play in Heim's (section 6); comment on another problem for the belief-set-based approach to want ascriptions first pointed out by Villalta (2008) (section 7); note how our view might be generalized to account for want ascriptions involving contextual alternatives as identified by Villalta (2008) and Lassiter (2011) (section 8); and say a few words about some ramifications that our proposal has for the broader landscape of desire predicates, beyond just 'want' (section 9).

6 Conditionals in Heim's semantics

It is a familiar thought that conditionals are intertwined with the semantics of 'want'. Indeed, the thought is central to Heim's approach to 'want'. She writes:

> The analysis of desire verbs I want to pursue here is sketched in Stalnaker (1984, p. 89): "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." An important feature of this analysis is that it sees *a hidden conditional* in every desire report. A little more explicitly,

> the leading intuition is that John wants you to leave means that John thinks that if you leave he will be in a more desirable world than if you don't leave. (Heim 1992, p. 39; emphasis ours)

We, of course, agree that there are hidden conditionals within 'want' reports; indeed, Heim's intuition is our major inspiration. To better understand how Heim's view relates to ours, it will help to locate the "hidden conditional" within her semantics.

Heim states her semantics in a dynamic framework. Stated in a static framework, and with our nomenclature, the semantics is this. Where $w <_S w'$ means that S prefers w to w':

(21) *Heim's semantics*[18]
$[\![\text{S wants p}]\!] = 1$ iff $\forall w \in \text{Bel}_S$: $Sim_w(\text{Bel}_S \cap p) <_S Sim_w(\text{Bel}_S \cap \neg p)$.

Put into words: ⌜S wants p⌝ is true just if, for every world w in S's belief set, S prefers the closest world in her belief set where p is true to the closest world in her belief set where $\neg p$ is true.

Now, let us identify just where the "hidden conditional" lies within this semantics. We can, as we did with von Fintel's semantics, factor Heim's semantics into three parts:

(5) *Form of a Semantics for 'Want'*
$[\![\text{S wants p}]\!] = 1$ iff S prefers the p worlds in D to the $\neg p$ worlds in D.

(6) *Belief Set Domain*
$D = \text{Bel}_S$.

(7) *Heimian Preference in a Domain*
S prefers the p worlds in D to the $\neg p$ worlds in D iff $\forall w \in D$:
$\text{Sim}_w(D \cap p) <_S Sim_w(D \cap \neg p)$.

On this way of viewing Heim's semantics, the "hidden conditional" is *in the notion of preference within a domain*. We, in contrast, propose a conditional *in the construction of the domain*. And our view is compatible with the Heimian preference within a domain (just as it's compatible with a best-worlds notion of preference in a domain).[19]

[18] Heim's Sim returns multiple worlds rather than just one (see footnote 15), and she uses not Bel but rather a superset of it that screens off intention-grounded beliefs (see footnote 6). These differences are not relevant here.

[19] How to adjudicate between best-worlds preference and Heimian preference? One consideration is that the two approaches make different predictions about the monotonicity of want ascriptions. Questions about monotonicity are well beyond the scope of this paper (see

7 Villalta's problem

Although we have solved the (C)FWA problem, one that vexed the belief-set approach, we inherit, and do not solve, another problem that the belief-set approach faces. This other problem, identified by Villalta (2008), is that the belief-set approach validates the following inference:

(22) a. *S* wants *p*.
b. *S* believes that *p* iff *q*.
c. ⇒ *S* wants *q*.

The inference in (22) runs counter to intuition. The following instance of (22), for example, is not good reasoning:

(23) a. Johnson wants to eat pizza.
b. Johnson believes that he'll get heartburn iff he eats pizza.
c. ⇒ Johnson wants to have heartburn.

If the want ascriptions in question are non-(C)FWAs, then our semantics faces the problem: it incorrectly validates the invalid inferences. This is because with non-(C)FWAs, our semantics is exactly the same as the belief-set-based semantics. In other words, the domain is the belief set.

Suppose, for example, that (23-a) and (23-c) are non-(C)FWAs. (Imagine that Johnson will eat pizza if it's available, but he's unsure if it's available, and so unsure if he'll eat it.) On our view, the domain for a non-(C)FWA is the agent's belief set. So, by (23-a), pizza worlds are best in Johnson's belief set; by (23-b), the pizza worlds in Johnson's belief set are heartburn worlds; so, heartburn worlds are best in Johnson's belief set, which is to say that (23-c) is true.

Happily, there are various solutions to this problem that can be combined with our view: the semantics of Crnič (2011), Phillips-Brown (2018, ms), and Dandelet (ms).[20] Each of these semantics says that the domain contains not

Crnič 2011 for discussion), and we do not want to take an official stance on this thorny issue. We can, however, point out one connection that our proposal bears to the debate. von Fintel (1999) argues that apparent cases of non-monotonicity in desire ascriptions are monotonic after all, once we take into account how the domain of a desire ascription shifts with the context. One way in which this can happen is when the agent's beliefs change: on the belief-set-based approach, this has a concomitant effect on the domain of the desire ascription. But not all apparent cases of non-monotonicity in desire ascriptions are attributable to changes in belief, and some such cases are argued by von Fintel (1999), building on prior work by Linebarger (1987), to involve implicit conditionalization. Our conditional-belief domain, when coupled with best-worlds preference, can be viewed as a continuation in this vein, fleshing out one way in which the domain can shift from one desire ascription to the next: moving from a (C)FWA to a non-(C)FWA or vice versa, even as the agent's belief set is held constant.

[20] As mentioned in section 3, Rubinstein (2017) also offers a solution to Villalta's problem,

possible worlds, but rather entities of some other, *coarser* kind: *situations* (see e.g. Kratzer 2019), *possibilities* (see e.g. Humberstone 1981), or *propositions*, depending on which of the authors you ask. We won't go into the details, but one may use our basic, conditional-belief recipe for constructing the domain with any of these views, substituting worlds (from our view) with any of these other types of entities.

8 Contextual alternatives: How to generalize our view

Our domain for ⌜S wants p⌝ is generated simply by the prejacent and its negation: it represents *S*'s conditional beliefs about *p* and about ¬*p*. Data from Villalta (2008) and Lassiter (2011), however, suggest the connection between the prejacent and its negation (on the one hand) and the domain (on the other) may not be quite so simple. In this section, we present Villalta's data and propose a generalization of our view that can accommodate it.

Here is a case structurally analogous to one of Villalta's (on her p. 469). Poe is the Secretary of State, and he prefers diplomacy to violence. The President, however, is a hawk. The country faces two enemies: the pretty bad guys and the really bad guys. Poe had proposed to the President three options, which in descending order of Poe's preference are: negotiate with the really bad guys, bomb the really bad guys, bomb the pretty bad guys. The President immediately dismisses the idea of negotiating with the really bad guys and says she'll soon decide between the remaining two options. Where all caps indicates emphasis:

(24) Poe wants to bomb the REALLY BAD GUYS.

(25) Poe wants to BOMB the really bad guys.

One can hear (24) as true and (25) as false. The reason for this is fairly intuitive: in (24), emphasis on *really bad guys* evokes a comparison to bombing the *pretty bad guys*, whereas in (25), emphasis on *bomb* evokes a comparison to *negotiating* with the really bad guys. Spelling this out a bit more, (24) is true because Poe prefers what he believes will happen if he bombs the *really bad guys* (loss of life) to a certain contextual alternative—what he believes will happen if he bombs the *pretty bad guys* (loss of life, but more gratuitous because the pretty bad guys are only pretty bad, rather than really bad). In contrast, (25) is false, one feels, because Poe *dis*prefers what he believes will happen if

one that involves giving want ascriptions an underspecified, context-dependent domain. For the reasons articulated in that section, however, we do not adopt this solution.

he *bombs* the really bad guys (loss of life) to a different contextual alternative than before—in this case, the alternative is what he believes will happen if he *negotiates* with the really bad guys (no loss of life).

Our blueprint for constructing the domain cannot straightforwardly account for this type of sensitivity to contextual alternatives. Our domain simply compares p to $\neg p$; there is no room for the type of shift in alternatives that Poe's case suggests. Fortunately, there is a straightforward way to generalize our blueprint that can account for this, a generalization that preserves our core insight that conditional belief is central to constructing the domain. We postulate—as Villalta, Levinson, and Anand and Hacquard (2013) do—a contextually determined set of propositions, A_c, against which ⌜S want p⌝ is evaluated in a context c. The domain needn't represent S's conditional beliefs about p and $\neg p$ in every context, but rather her conditional beliefs about each member of A_c (which in certain contexts may well include only p and $\neg p$). Specifically, where D_c is the domain in c:

(26) *Generalized Conditional Belief Domain*
$D_c = \bigcup\{\text{ConBel}_S(q) \mid q \in A_c\}$

This account can capture the intuitive diagnosis of Poe's case. In the context c where (24) is true, A_c is {*bomb the really bad guys, bomb the pretty bad guys*}, and so D_c is $\bigcup\{\text{ConBel}_{Poe}(\textit{bomb the really bad guys}), \text{ConBel}_{Poe}(\textit{bomb the pretty bad guys})\}$. This is to say that the domain represents what Poe believes will happen if he bombs the really bad guys and what he believes will happen if he bombs the pretty bad guys—exactly what we were looking for. In the context c' where (25) is false $A_{c'}$ is {*bomb the really bad guys, negotiate with the pretty bad guys*}, and so $D_{c'}$ is $\bigcup\{\text{ConBel}_{Poe}(\textit{bomb the really bad guys}), \text{ConBel}_{Poe}(\textit{negotiate with the really bad guys})\}$. In other words, the domain represents what Poe believes will happen if he bombs the really bad guys and what he believes will happen if he negotiates with them—again, exactly what we were looking for.[21]

This is how we can account for Villalta's alternative-sensitive data (and Levinson's too, although we won't go into that here). We can account for all of the data we've encountered previously, too, because when $A_c = \{p, \neg p\}$, the

[21] Ultimately, a comprehensive treatment would need to show how $A_c/A_{c'}$ is determined not just by the discourse context but also by the intonational emphasis in the want ascription being evaluated. We do not undertake this here, because it would take us too far afield, but the point to be stressed is that the kind of pattern we see here instantiates a much broader linguistic phenomenon known as focus sensitivity (see Villalta 2008 for discussion and references), so what we are suggesting here recycles technology independently needed elsewhere in the grammar.

new, generalized domain, $\bigcup\{\text{ConBel}_S(p), \text{ConBel}_S(\neg p)\}$, is exactly the same set as our previous domain, $\text{ConBel}_S(p) \cup \text{ConBel}_S(\neg p)$. We'd like to tentatively propose—following Anand and Hacquard (2013), whose semantics is alternative-sensitive, albeit within a different framework—that the default for A_c is $\{p, \neg p\}$, and so the default domain is our previous domain.

In the next section we will, for ease of exposition, return to discussing the previous domain, but what we say applies just as well to this generalized domain.

9 The landscape of desire predicates

What is the analytical relationship between 'want' and its cousins 'wish', 'be glad (that)', and 'hope'? Our proposal enables the view that all four of these predicates share the same core semantics, including the conditional belief domain:

(27) *Form of a semantics for 'wants'/'hopes'/'wishes'/'is glad that'*
$[\![\text{S } \{\text{wants/hopes/wishes/is glad that}\} \text{ p}]\!] = 1$ iff S prefers the p worlds in D to the $\neg p$ worlds in D.

(28) *Conditional belief domain for 'wants/'hopes'/'wishes'/'is glad that'*
$D = \text{ConBel}_S(p) \cup \text{ConBel}_S(\neg p)$
$= \{w' \mid \exists w \in \text{Bel}_S\colon \text{Sim}_w(p) = w'\} \cup \{w' \mid \exists w \in \text{Bel}_S\colon \text{Sim}_w(\neg p) = w'\}$

(29) *Best-worlds preference in a domain for 'wants'/'hopes'/'wishes'/'is glad that'*
S prefers the p worlds in D to the $\neg p$ worlds in D iff $\forall w \in \text{best}_S(D)$: $p(w) = 1$.

What distinguishes these four desire predicates from each other (or at least, one crucial dimension along which they are distinguished) are the presuppositions that they impose on the relationship between p and S's belief set. 'Want' has no such presupposition, whereas—in line with much previous literature—'wish' presupposes that S does not believe p, 'be glad' presupposes that S does believe p, and 'hope' presupposes that S neither believes p nor believes $\neg p$:

(30) *Doxastic presuppositions of desire predicates*

'want'	none	
'wish'	S believes $\neg p$	$\text{Bel}_S \subseteq \neg p$
'be glad'	S believes p	$\text{Bel}_S \subseteq p$
'hope'	S neither believes p nor $\neg p$	$\text{Bel}_S \not\subseteq p \wedge \text{Bel}_S \not\subseteq \neg p$

These doxastic presuppositions interact with our uniform Conditional Belief Domain to ensure that the domain for 'hope' ends up simply being S's belief set, while the domain for 'wish' ends up being S's belief set expanded to include maximally similar p worlds, and the domain for 'be glad' ends up being S's belief set expanded to include maximally similar $\neg p$ worlds.

That the domain for 'hope', unlike 'want', is S's belief set and that 'hope' but not 'want' requires the subject to believe the desired proposition to be possible finds support in such previous work as Portner (1992); Portner and Rubinstein (2012); Anand and Hacquard (2013); Silk (2018); Blumberg (ms). Compare, for example, Heim's (3), repeated below, and a variant with 'hope' in place of 'want':

(3) I **want** this weekend to last forever (but of course I know it will be over in a few hours).

(31) ??I **hope** this weekend will last forever (but of course I know it will be over in a few hours).

As we know, (3) does *not* to commit the speaker to belief in the possibility of a never-ending weekend; she believes that the weekend will soon be over. In contrast, (31) is infelicitous exactly because the 'hope' ascription *does* seem to commit the speaker to the possibility of a never-ending weekend, which is what she disavows in the parenthetical.

Moreover, our domains for 'wish' and 'be glad' end up being essentially the same as what Heim (1992) proposes for these two predicates, though we improve on her approach in that we derive these domains from the interaction between their doxastic presuppositions and a uniform domain rather than stipulating these domains lexically for each predicate.

Our rethinking of the boundaries between these four desire predicates does raise an important question. It is well documented that when two lexical items differ only in such a way that one is presuppositionally stronger, the presuppositionally weaker member typically comes across as infelicitous in contexts where the stronger presupposition is satisfied—this is Heim's (1991) Maximize Presupposition. Observe, for example, the following sentences. In each case,

choosing the first word in the disjunct sounds odd because there is an alternative word that better satisfies uncontroversial background assumptions (that there is only one sun, that I have exactly two eyes, and that 2+2=4, respectively).

(32) *Examples of Maximize Presupposition*

a. {??**A**/**The**} sun is shining.

b. {??**All**/**Both**} of my eyes are closed.

c. John {??**believes**/**knows**} that 2+2=4.

Why, then, is 'want' possible even in contexts where the presuppositions of 'wish', 'be glad', or 'hope' would be satisfied? We begin by noting that under some conditions, it may be possible to detect a kind of weak Maximize Presupposition effect. Blumberg (ms), for example, says that it would be odd to say 'Bill knows that Mary has a terminal illness. He **wants** her to get better.' (compare: '…He **wishes** she could get better.'), and even invokes Maximize Presupposition to explain why it is odd. But the effect, if there is one, is rather weak, and we suspect that it may reflect a stylistic preference (think of a careful writer wanting to choose the most appropriate, most precise word) rather than being the business of grammar. And we furthermore suspect that the reason for this is that there are additional semantic differences distinguishing 'want' from its presuppositionally stronger counterparts that prevent genuine Maximize Presupposition competition.

While it will be beyond the scope of this paper to explore these semantic differences in detail, we can offer some suggestive data that illustrate some of them. As for 'want' vs. 'hope', Portner and Rubinstein (2012) note, for example, the contrast in acceptability between the two predicates in

(33) He doesn't fully realize it yet, but Ron **wants/??hopes** to date Hermione. (p. 471)

which persists even on the assumption that Ron believes dating Hermione to be a possibility. Portner and Rubinstein's own proposal is that 'hope' but not 'want' involves the agent's contextual commitment to a preference, one requirement of which is the agent's own awareness of the preference. Regardless of whether we adopt this particular approach, though, the crucial point is that 'want' differs from 'hope' in ways that prevent Maximize Presupposition competition.

As for 'want' vs. 'be glad', we note the following FWA due to Iatridou:

(34) I live in Bolivia because I **want** to live in Bolivia. (Iatridou 2000:243)

Observe that its 'be glad' counterpart sounds quite odd:

(35) ??I live in Bolivia because I **am glad** that I live in Bolivia.

The same contrast is apparent in comparing Heim's (1992) FWA (see footnote 6 above) with its 'be glad' variant:

(36) John hired a babysitter because he {**wants** to go/??**is glad** that he'll go} to the movies tonight.

For reasons that are not entirely clear to us, it appears that states of wanting can be used in *because*-clauses to explain certain kinds of behavior (living in Bolivia, hiring a babysitter) in ways that states of being glad cannot. Whatever the reason is, it points to some semantic difference between 'want' and 'be glad'—beyond what they presuppose or not about the agent's belief set— that plausibly blocks Maximize Presupposition competition.

Finally, what about 'want' vs. 'wish'? We are not aware of any CFWA that cannot be felicitously paraphrased with a wish ascription. Take, for example, (37), just below, which can be felicitously paraphrased by its 'wish' counterpart, (38):

(37) I **want** this weekend to last forever.

(38) I **wish** this weekend would last forever.

But in general the reverse is not true. Specifically, wish ascriptions can be used for preferences about past counterfactual scenarios, for example:

(39) 'I **wish** I had been there yesterday.'

Here in (39), a want paraphrase sounds rather awkward:

(40) '??I **want** to have been there yesterday.'

This suggests that 'want' but not 'wish' comes with the requirement that the desired proposition must be temporally simultaneous or future-oriented with respect to the time of the desire. And we suggest that this difference—or whatever more fundamental difference might be responsible for it—may account for the lack of Maximize Presupposition competition between 'want' and 'wish'.

10 Conclusion

The goal of this paper was to revise the belief-set approach to want ascriptions in a way that solves the well-known problem of (C)FWAs, while doing as little damage as possible to all of the virtues of the belief-set approach. Indeed, the persistence of the belief-set approach in spite of its well-known shortcoming is a testament to the many results that it does get right, and we offer the proposal in this paper as one among many contributions leading to a more sophisticated, more accurate version of the belief-set approach.

In carrying out our goal, we have shown that conditional belief—and more specifically a model of conditional belief inspired by Stalnaker's approach to conditional sentences—offers an elegant way of unifying non-(C)FWAs, CFWAs, and FWAs under a single semantics that strikes the right balance between rigidity and flexibility in characterizing the domain.

Finally, we have reflected on the consequences of our proposal for rethinking the boundary between 'want' in relation to its cousins 'hope', 'wish', and 'be glad', in ways that contribute to increasingly fine-grained and accurate lexical semantics for desire predicates.

References

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

Arregui, A. (2007). When aspect matters: the case of would-conditionals. *Natural Language Semantics*, 15(3):221–264.

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

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

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.

Edgington, D. (2014). Indicative conditionals. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2014 edition.

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.

Gardenfors, P. (1982). Imaging and conditionalization. *The Journal of Philosophy*, 79:747–760.

Geurts, B. (1998). Presuppositions and anaphors in attitude contexts. *Linguistics and Philosophy*, 21(6):545–601.

Heim, I. (1991). Artikel und definitheit. In von Stechow, A. and Wunderlich, D., editors, *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, pages 487–535. de Gruyter, Berlin.

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

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

Iatridou, S. (2000). The grammatical ingredients of counterfactuality. *Linguistic Inquiry*, 31(2):231–270.

Ippolito, M. (2013). *Subjunctive Conditionals: A Linguistic Analysis*. The MIT Press.

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.

Kratzer, A. (2019). Situations in natural language semantics. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*.
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. (1973). *Counterfactuals*. Blackwell.
Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *The Philosophical Review*, 85:297–315.
Linebarger, M. C. (1987). Negative polarity and grammatical representation. *Linguistics and Philosophy*, 10:325–387.
Phillips-Brown, M. (2018). I want to, but.... In Truswell, R., Cummins, C., Heycock, C., Rabern, B., and Rohde, H., editors, *Sinn und Bedeutung 21*, pages 951–968.
Phillips-Brown, M. (fc). What does decision theory have to do with wanting? *Mind*. Forthcoming.
Phillips-Brown, M. (ms). Some-things-considered desire. Unpublished manuscript.
Portner, P. (1992). *Situation theory and the semantics of propositional expressions*. PhD thesis, University of Massachusetts: Amherst.
Portner, P. and Rubinstein, A. (2012). Mood and contextual commitment. *Semantics and Linguistic Theory 22*, pages 461–487.
Rubinstein, A. (2012). *Roots of Modality*. PhD thesis, University of Massachusetts Amherst.
Rubinstein, A. (2017). Straddling the line between attitude verbs and necessity modals. In *Modality Across Syntactic Categories*, pages 610–633. Oxford University Press.
Scheffler, T. (2008). *Semantic operators in different dimensions*. PhD thesis, University of Pennsylvania.
Silk, A. (2018). Commitment and states of mind with mood and modality. *Natural Language Semantics*, 26:125–166.
Stalnaker, R. (1968). A theory of conditionals. In Rescher, N., editor, *Studies in Logical Theory (American Philosophical Quarterly Monographs 2)*, pages 98–112. Blackwell.
Stalnaker, R. (1975). Indicative conditionals. *Philosophia*, 5(3):269–286.
Stalnaker, R. (1984). *Inquiry*. Cambridge University Press.
Villalta, E. (2008). Mood and gradability: An investigation of the subjunctive mood in Spanish. *Linguistics and Philosophy*, 31(4):467–522.
Wrenn, C. (2010). A puzzle about desire. *Erkenntnis*, 73(2):185–209.