The Attitudes We Can Have*

Daniel Drucker

Keywords: Belief, Attitudes, Choice, Questions, Rationality

In this paper I'm going to explain why there are many propositions that we have good reason to believe are true but that we cannot rationally believe, and why this kind of phenomenon only occurs with doxastic attitudes like belief and not non-doxastic attitudes like admiration. To do that I will present a general theory of attitude formation as a kind of choice.

My focus will be on *options*. Our options frame our choices, and because different option sets make the same action rational in one case and irrational in another, the frame determines what we may rationally do; and if it is more difficult to do what is manifestly irrational, they determine what we *can* do. That is the conventional wisdom when we have in mind "ordinary" choices like where to go to graduate school, what to eat for dinner, and which dog to get. It helps to think of other things we do this way, too, for example coming to believe what we believe or admire whom we admire. Or so I will argue.

That argument will have two components. First, seeing attitude formation as choices among options will allow us to see our way through a tangle of puzzles for belief and the non-doxastic attitudes; I will focus on admiration, hope, hatred, and anger. Second, I'll argue that this is independently a natural way of thinking about attitude formation.[1] In sections 1 and 2 I will present those puzzles. In section 3, I will show how a similar puzzle arises for choice, and I will present a requirement of rationality, Clarity, that solves the puzzle for choice. In section 4, I will argue that belief formation ought to be conceived of as a choice among options for independent reasons, and in section 5 I will argue that doing so allows us to solve the puzzles I presented in the first section. In section 6, I will argue that we also have independent reasons for thinking of non-doxastic attitude formation as choices,

*I've gotten a lot of help with this paper from a lot of friends. Thank you to Axel Barceló, Maria Lasonen-Aarnio, Ricardo Mena, and Sarah Moss, and audiences at UNAM and the University of Helsinki for comments and discussion. Two referees and the editors of this journal provided incredibly helpful comments that changed, and I hope improved, the paper enormously. Dmitri Gallow read maybe five different versions and offered fundamental helpful comments and corrections every single time. Oscar Zoletto went through the entire thing with me, word-by-word. And Eric Swanson helped me, and showed faith in me, from beginning to end.

[1] Recently a more decision-theoretic view of belief and credence has become popular. See, e.g., Joyce (1998) and Greaves (2013). This viewpoint is congenial to mine, but I will not presuppose it.

too, and that doing so helps solve the puzzles I raised for them. Finally, in the appendix, I discuss some interesting conditionals that present a (merely) *prima facie* problem for the way I set up the puzzles in sections 1 and 2.

1 The Problem of Impossible Modally Unstable Belief

In this section and the next, I will present a puzzle involving belief and some other attitudes. Most of the work I do in these sections will be formulating generalizations whose truth I will go on to explain in later sections.

It is as far as I can tell unknown whether there was a single person, Homer, who uniquely wrote both the *Iliad* and the *Odyssey*. Call the proposition that there was a single person, Homer, who uniquely wrote both the *Iliad* and the *Odyssey* 'Homerism'. This is just like calling the proposition that mathematics reduces to logic 'logicism'; it seems to me that we can and do name propositions this way.[2] Call the negation of Homerism 'anti-Homerism'. If we assume bivalence, one of Homerism and anti-Homerism is true, and the other false. Let '*H*' name the true member of the set {Homerism, anti-Homerism}. I will suppose from now on that anti-Homerism is true, and therefore that '*H*' names anti-Homerism. All sides should be agreed I have strong evidence that *H* is true, since my evidence obviously *entails* that *H* is true. But, though perhaps I believe that *H* is true, I don't believe *H* itself, i.e., anti-Homerism—I'm agnostic about whether Homer really existed, and given my overall epistemic circumstances, I should be.[3]

That's because, on ordinary uses of the word 'evidence', this seems clear: I don't have strong evidence in favor of *H*, i.e., of the proposition that there wasn't a unique individual, Homer, who wrote the *Iliad* and the *Odyssey*. I know roughly what strong evidence for that proposition would be. It'd be finding, perhaps, burial inscriptions that we can date to the right period that ascribe these accomplishments to the person buried. I don't have anything like that kind of evidence. And I certainly don't get it with any naming tricks. But I *do* have strong evidence that *H* is true, because my evidence entails that '*H* is true' is true, given my awareness of the stipulation I made; and I am also certain that *H* is true iff '*H* is true' is true, based on my linguistic competence. If that's right, I have strong evidence that *H* is true, while only weak evidence for *H* itself. Assuming a weak kind of evidentialism, I can rationally believe that *H* is true in my circumstances, even if I can't rationally believe *H* itself, i.e., that there wasn't a unique individual, Homer, who wrote the *Iliad* and the *Odyssey*. Were I to believe *H*, I would only be able to by a wild guess, and that would

[2] For discussion, see, among many others, King (2002), and a bit later in this section.

[3] In this paper I will assume that belief is a propositional attitude, relating believing subjects to propositions. I will also assume that believing that *p* and believing the proposition that *p* are the same thing. This can be questioned: Moltmann (2003), and more recently, Nebel (2019) reject some of these assumptions; but these assumptions won't play anything beyond a simplifying role in my arguments.

be irrational. That's one part of my puzzle, that believing that H is true is rational in my circumstances but believing H itself isn't.[4]

Let's be a little more careful about what, exactly, that part of the puzzle is supposed to be.[5] The case of 'H' raises the following questions:

- Is the 'H'-stipulation so much as possible?
- Is it rational to believe that H is true?
- Is it irrational to believe H?

If we answer "yes" to all these questions, as I have suggested we should, then the phenomenon of interest follows: it can be rational to believe that a proposition is true even though it's not rational to believe that proposition itself. That would call out for an explanation itself—it would be puzzling—and it would be pressing whether, in explaining it, we have to give up any part of our commonsense picture of belief. By 'commonsense picture of belief', I just mean ordinary, pre-theoretical conceptions of when we'd ascribe beliefs to people and when we'd withhold those ascriptions; and I also mean our sense of the kinds of evidence that would support belief in those propositions if we possessed it and of the kinds of evidence that would not. Ultimately, I will present an account that vindicates the "yes" answers to these questions without giving up any part of the commonsense picture of belief. That doesn't mean there wasn't a puzzle here, of course; it just means that I think we can satisfyingly solve that puzzle.

You might prefer to answer "no" to one of those questions. If you did, that would be yet another way to solve the puzzle. But to show that it still would have been a puzzle, I'll argue that answering "no" to one of these questions involves theoretical costs we are not otherwise forced to incur. So, for example, you might deny that the stipulation is so much as possible. I accepted the possibility of such stipulations because, for example, I think that we can name pretty much *any* object *via* such stipulations. There will be weird cases ("let 'L' name the tallest person eternally unnamed in my idiolect"), but on the whole, it seems like

[4] This calls to mind two kinds of puzzle, but mine is different from both. First, that we cannot believe "at will" (see, e.g., Williams (1973) and Hieronymi (2006), among many others). It's related but different because, though this isn't about voluntarism, in that literature belief needs to be subjectively rationalized for the subject. Second, there's the puzzle presented in Kaplan (1968), Kripke (1972), and Donnellan (1977). They use descriptions and reference-fixing tricks to raise puzzles about why certain belief ascriptions are false (or appear to be). There are big differences between my projects and theirs, though. First, I am interested in the objects of attitudes, rather than the objects the given beliefs are *about*. Second, I will look at *non-doxastic* attitudes for which the analogue of this sort of belief are *possible*. Third, one can accept that I cannot rationally believe H in my circumstances, but still accept (or reject) the following: I believe *of* the shortest spy that they are a spy and I know that Julius invented the zipper (see a bit further down for the second example). I am focused on *de dicto* belief, not *de re*.

[5] I thank a referee for the impetus to get clearer on what I think this part of the puzzle is and why standard frameworks don't already solve it.

we can name any *kind* of object. But certain kinds of Fregean might deny the possibility of such stipulations, since they might have an extraordinarily fine-grained view of the objects of belief according to which any possible difference in rational attitude toward p and p' entails that $p \neq p'$. (I say "certain kinds" because many more minimal Fregeans are not committed to denying this possibility, e.g., those who think we can refer to senses and be unsure whether this sense is the same sense as that sense, etc.[6]) But it would be strange if we couldn't come up with names for objects of belief, and so I think this way of responding to the puzzle about 'H' incurs a theoretical cost.

I will assume that, once we agree that the stipulation is possible, no one will deny that it's rational to believe that H is true; we need only be aware of the stipulation. Instead, you may answer "no" to the third question, claiming that we *can* rationally believe H, i.e., that there wasn't wasn't a unique individual, Homer, who wrote the *Iliad* and the *Odyssey*. Other kinds of Fregean, perhaps those who think there are just two objects of belief (the True and the False), might accept this. That is an extreme view and not one it is easy to endorse, since on the commonsense picture, there are some true objects of belief some people don't believe. Next, some Millians endorse certain belief ascriptions that someone who endorsed the commonsense picture of belief would reject. They think, e.g., that 'Lois believes Superman can fly' and 'Superman = Clark Kent' jointly entail 'Lois believes that Clark Kent can fly'.[7] But that they endorse this entailment, this departure from the commonsense picture of belief, is a theoretical cost, one that they make up elsewhere by, e.g., being able to endorse an attractive semantic theory of proper names—direct reference—and some general semantic principles (e.g., semantic innocence). To deny that it is irrational for someone in my circumstances to believe H would be to incur *additional* theoretical costs beyond what the Millian is already committed to incurring. The Millian position doesn't say anything about whole propositions like H; I didn't appeal to anything about the semantics of proper names or semantic innocence. Further, the commonsense picture includes with it a sense of what counts as evidence in favor of believing what. On this picture, good evidence for H, as I said, would be like the burial inscriptions I mentioned before. If you were to ask classicists, e.g., that's the sort of evidence they'd mention and seek. Notice also that we can go through the same kind of stipulation procedure with any other proposition that I can grasp that I did with 'H'; it's fully general. Thus the person who answers "no" to the third question is committed to the possibility of a person's rationally believing *any* true proposition that they can grasp or even just describe without having any of the canonical kinds of evidence for those propositions. So the person who answers "no" to the third question

[6] Some Fregeans that go as far as Frege does in denying that we can refer to the properties we predicate explicitly allow for reference to propositions ("information units") in their theory. See, e.g., Chierchia and Turner (1988). It's rare to see Fregeans say we can't refer to objects of belief.

[7] See, e.g., Salmon (1986) and Braun (1998).

would incur large theoretical costs beyond what just the Millian incurs. Perhaps these costs are worth paying. Or perhaps not—if we didn't have to pay anything comparable, they wouldn't be. And one last benefit of my solution is that I will be better positioned to explain some intuitive linguistic data than someone who thinks we do believe *H* and anything else we stipulate that way.

So one way to think about this part of the puzzle is: are we forced by '*H*'-like cases to deny more of the commonsense picture of belief than it looked like we had to deny at first? My answer will be "no"; but then I have to explain how rationally believing *p* and rationally believing that *p* is true can come apart. That's what I will do in later sections of this paper. Others will answer "yes". Which of us is right constitutes this part of the puzzle generated by '*H*'.

So, suppose we do answer "yes" to the three questions I listed. Why should that be puzzling? Importantly, it suggests (in fact, seems to presuppose) that the proposition that *p* is a different proposition from the proposition that *p* is true. But many philosophers would have denied that, for example philosophers who think propositions are sets of possible worlds or those who think 'true' doesn't denote a substantial property of propositions (but is rather an anaphoric or summarizing device). That said, I take myself to have a *prima facie* plausible argument against identifying the two kinds of propositions: in my circumstances, given how I baptized '*H*', it is rational for me to believe that *H* is true but, as I just argued, it is not rational for me to believe that *H*. So by Leibniz's Law they are distinct. Others have other reasons for rejecting this identification.[8] But I won't argue for it much further;[9] I will just assume that it is right, and then explain how it can be rational for me to believe that *H* is true but not to believe *H*.

The problem I've been developing is just an instance of a broader phenomenon, but its exact contours require care in describing. An individual in my circumstances cannot believe *H* rationally. Why not, though? I know that were I to believe *H*, I would believe a truth, and an interesting one at that, since either proposition in {Homerism, anti-Homerism}

[8] Fans of structured propositions likely fall in this camp, but also Jerzak (2019) argues for this in connection with the paradox of the knower. A referee suggests that we might use Mates (1952)-style cases to argue for this conclusion, too. Take, e.g., someone who thinks 'true' means 'false'. Then (the referee suggests) they might believe it's true that snow is green without believing that snow is green, and even that the intuitions are clearer in this case than in mine. But I'll just note my disagreement here. I think that, though Burge has shown us that we may have incomplete mastery of a concept but still employ the concept in believing as we do, such dramatic misunderstandings as in the referee's case suggests that the person really believes it's *false* that snow is green, and they use 'true' in their language of thought to encode that belief. Of course, matters here are very difficult, but at least I don't think the intuitions are clearer in that case than in mine.

[9] You may worry, as a referee does, that my Leibniz's Law argument depends too much on intuitions about descriptive names, and that intuitions about such names are just too dubious to place much weight on. But actually, it doesn't. We can use, e.g., demonstratives ('*that* proposition that is the true element of {Homerism, anti-Homerism}, e.g.). All I need is a way of referring to these propositions *via* that description, and once I'm granted that, everything proceeds in the same way. '*H*' makes the point simply, but there are more ways than that.

would be interesting to know. I'll now go through a couple of suggestions about what the phenomenon is and how it might arise.

First, maybe I can only believe H if I know which proposition H is. This might be right, though I have my doubts.[10] Mostly I think the notion of knowledge-'wh' (here, knowledge-'which') is too context-sensitive in ways that still remain pretty obscure to do much philosophical work with right now.[11] I will sometimes talk about knowing which proposition something is, specifically in section 5, but in a circumscribed sense clearly equivalent to a particular knowledge-'that' claim.

Second, you might think the phenomenon arises from how I fixed 'H"s reference; perhaps the problem is that I used the word 'true'. It's true that that allowed me to claim that we (I, you, anyone else) have great evidence that H is true, but other reference-fixing descriptions would allow me to make similar claims. I don't know whether neanderthals were on average smarter than humans. Let 'Neanderthalism' denote the proposition that they were smarter on average, and 'anti-Neanderthalism' denote the proposition that they were not. Finally, let 'N' denote the proposition in the set {Neanderthalism, anti-Neanderthalism} that the majority of biological anthropologists believe (making the simplifying assumption here that there's no exact tie). Once again, I have strong evidence that N is true, but I can't believe it, at least not rationally; and if it's obvious that I can't believe it rationally, then I typically can't believe it at all. 'True' isn't essential to motivate the puzzle.

The third possible characterization is one that I went through earlier, that the phenomenon of interest occurs when a person's evidence supports the proposition that a proposition is true without supporting that proposition itself. This characterization maintains what's puzzling about the phenomenon. After all, we can typically infer a proposition p from the proposition that p is true, and *vice versa*. That is, we can typically rationally come to believe the one on the other's basis and *vice versa*. So *prima facie* we ought to believe both p and the proposition that p is true to the same degree, and thus we should expect that the

[10] They concern the following sorts of case, adapted from Soames (1989, page 193).

Logicism. Let 'logicism' denote the proposition that mathematics is reducible to logic. A student in an introductory philosophy of mathematics class might know that logicism is *some* proposition about the relation between mathematics and logic, and they might even know things like "Frege was an important philosopher who believed that logicism was true" or "if logicism is true, then formalism is false". But for them, logicism might even *be* formalism!

Does the student know what proposition "if logicism is true, then formalism is false" expresses? I can imagine going either way. After all, for all they know, 'logicism' expresses the proposition that 'formalism' expresses. I feel torn on what to say of cases like these, which is why I have doubts, rather than positively disbelieve the proposal.

[11] In an interestingly similar dialectical context, Quine (1981) says: "The notion of knowing or believing who or what someone or something is, is utterly dependent on context. [...] Of itself the notion is empty." For a modern form of contextualism I find plausible, see Aloni and Jacinto (2014); and for interest-relativity, see Boër and Lycan (1975).

evidence supports them equally. But with H and N I've argued that that's not true. That means we haven't made progress on *explaining* the phenomenon, but here I only want to correctly and informatively *characterize* the phenomenon, not to explain it; that will come in later sections. But I still won't work with this characterization very much, because ultimately I want to draw a contrast between the doxastic attitudes and some non-doxastic ones using this behavior, but evidential support really only applies to the doxastic attitudes. So I would like a more general characterization of these cases that doesn't directly appeal to evidential support.

So, the first suggestion might be right, though too imprecise to do much philosophical work right now, while the second is wrong, and the third is too specific to doxastic attitudes. Here is the characterization of the phenomenon I prefer. Consider two people, S_1 and S_2. S_1 comes to believe N by surveying anthropologists about Neanderthalsm and anti-Neanderthalism, tallying their views, and believing what they find the majority believed. S_2, by contrast, simply believes it—perhaps *per impossibile*, but bracket that—by believing *whatever* the majority of anthropologists believe of Neanderthalism and anti-Neanderthalism. S_1 and S_2, whatever their similarities, would still be somehow different in their underlying cognitive states. To see this, note that S_2 could believe $\neg N$ were the majority of anthropologists to believe the negation of what they actually believe, even though S_2 remains internally identical in all relevant respects to how they actually are. For S_1 to change to believing $\neg N$, by contrast, they would have to change internally somehow. S_2's belief state would track majority expert opinion across modal space, without their exerting any specific effort in each world to do so in a way particular to that world, whereas S_1's would not. More generally, *S modally tracks F* things with her attitude state Ψ across a set W of worlds just in case S's internal duplicates in each world w in W bears Ψ to whatever's F in w.

In that definition, you may have noticed that I distinguished between beliefs and belief states, and more generally token bearings of Ψ to objects and underlying Ψ states. I take it that token attitudes are *at least* individuated by their contents. That is, if my token believing relates me to p, then—as I am using these terms—any belief I might have or might have had toward $q \neq p$ is not the same token belief. But I might have the same belief *state*.[12] Take the following example. Following Evans (1982), let 'Julius"s reference be fixed with the description 'the inventor of the zipper'. I say to myself: 'Julius is clever', thereby expressing my belief that Julius, i.e., Whitcomb L. Judson, is clever.[13] In another world where someone other than Whitcomb L. Judson invented the zipper, Florence Nightingale, say, I would

[12] I roughly mean this notion to track how its used in, e.g., Perry (1980), among other places.

[13] Here I assume direct reference theory for proper names (see the next paragraph for a brief explanation), even descriptive names like 'Julius'. If direct reference theory is wrong, my discussion could be simpler than it is here, but it wouldn't be adversely affected.

rather express the belief that Florence Nightingale was clever. These are different beliefs, at least as I would like to individuate them. But I want to say that this is the same belief *state*—the same underlying thing in my head, roughly. The same considerations apply, *mutatis mutandis*, to other kinds of attitude Ψ.

Perhaps non-constant doxastic modal tracking is possible, where the same belief state has different propositional objects in different worlds. The most plausible cases involve what we might call externalist content-determining mechanisms. Any examples I give will be at least somewhat controversial, but likely included are the mechanisms that assign contents to thoughts with mental indexicals and demonstratives, natural-kind concepts like WATER, and descriptive name concepts. Focus on the last sort for a moment. Continuing the earlier example, suppose I believe that Julius is clever, and as it happens X invented the zipper. Then my belief will be that X is clever, at least if names are directly referential. But consider an alternative world in which Y rather than X invented the zipper, but where mostly everything else is as it actually is, so that I'm relevantly internally identical. Now my belief will be that Y is clever. I thus non-constantly modally track a certain property of propositions, namely being a proposition that attributes cleverness to that person who in that world invented the zipper, at least assuming as we are for the moment that the propositions so tracked are different.

Still, even if non-constant doxastic modal tracking is possible, it is so only in relatively limited circumstances. Here is a generalization that avoids 'Julius'-type counterexamples:

> STABILITY CONSTRAINT FOR BELIEF. Belief states cannot (rationally) modally track propositional properties across a set of worlds W when different propositions have those properties in different worlds in W, unless they track them in that way because of some relevant externalist content-determining mechanism.

There are some things to note about this generalization. First, and most importantly, I have not yet *explained* it; doing that, and thereby also explaining why we can't believe H, is one of the overall aims of the paper. Second, it does not commit me to thinking that there *are* any externalist content-determining mechanisms. That is another controversy I need not enter into. But if there are, this principle accommodates them without being falsified or trivialized. That said, 'H' and 'N' do look structurally quite similar to 'Julius'—they are all names with their reference fixed by descriptions. In order for STABILITY CONSTRAINT FOR BELIEF to be a good generalization of the phenomena with which I began, then, propositional designators can't trigger the externalist exception.[14] Assuming content internalism,

[14] By this I mean: even if proposition names like 'H' do involve externalist mechanisms, there might still be some reason they don't enable doxastic modal tracking. I'll present this reason in section 5. Thanks to a referee here.

my explanatory burden is to explain why doxastic modal tracking is not (rationally) possible. But assuming externalism: why are there no externalist content-determining mechanisms that generate rational H- and N-beliefs? That is what I'll explain in sections 4, 5, and 6, along with the Stability Constraint for Belief itself.

Even before attempting to explain it, though, it will help to see more evidence *that* this generalization is correct. I have largely been arguing in the material mode, but many of the same considerations translate into the formal mode. Most simply, when the circumstances are as I've assumed, the following sound bad:

(1) #I believe whichever of Homerism and anti-Homerism is true.[15]

(2) #I believe whatever the majority of the experts believe about whether neanderthals were smarter than humans.

That's because the speakers don't, I'm assuming, have the relevant beliefs by digging through the evidence or surveying anthropologists.

In fact, the formal mode offers a relatively direct way of verifying that Stability Constraint for Belief is true. Consider:

(3) #If there was a unique individual who wrote the *Iliad* and the *Odyssey*, then I believe there was, and if not, then I believe there wasn't.

(4) ??If a majority of the experts believe that neanderthals were smarter than humans, then I believe neanderthals were smarter than humans, and if a majority believes not, then I believe they weren't.

(3) is simply not true when said by any speaker in anything like the relevant circumstances. (4) sounds bad, too, but somewhat better, I think. Actually we should *want* to say that (4) is false. If it were true then in a situation where the majority of the experts really did believe that neanderthals were smarter than humans, then a third party could use *modus ponens* to detach the consequent. In other words, conditionals like those expressed by (3) and (4) entail propositions like those expressed by (1) and (2), at least when the conjunctions of conditionals mention all the propositions that have the relevant property.

(This argument in particular is inconclusive because we might not want to understand (3) and (4) as conjunctions of conditionals of the form $\ulcorner$(if σ_1, I believe that σ'_1)$\wedge$ ... $\wedge$ (if σ_n, I believe that σ'_n)$\urcorner$. I'll discuss this more in section 2 and the appendix, but for now I'll just call the argument for (3) and (4)'s badness *prima facie*.)

[15] Key: Strings of '?' indicate dubiety of felicity, the more the more dubious; '#' indicates clear infelicity (a kind of least-upper bound of '?'s); and '*' indicates semantic or syntactic ill-formedness.

More generally, suppose the propositions that might be F are p_1, ..., p_n. Then someone who rationally believes whatever is F is such that they rationally believe p_1 if p_1 is F, p_2 if p_2 is F, ..., and p_n if p_n is F. This *may* be possible. Consider the following example:

(5) (*After making the 'Julius' stipulation*) I believe that Julius was clever. So, if Julius was Archibald Roverhamptonshire, then I believe he was clever, and if Julius was Benjamin Disraeli, then I believe he was clever.

If you have internalist intuitions, you will likely find (5) bad, but if you have externalist intuitions, you may well find it good. This is exactly what the Stability Constraint for Belief predicts.

Stability Constraint for Belief is meant simply to be a (quasi-)empirical generalization. It isn't self-evident what the best explanation for its truth is, even if there is strong evidence that it is true. One of my aims in this paper is to give an explanation of why it's true, and also of why I cannot rationally believe H, and I will do that in later sections. But first, in the next section I'll show why the problem is even harder than it seems, by showing that analogues of the Stability Constraint for Belief do not apply to many other kinds of attitude.

2 The Contrast with Belief, and the Problem of Outsourced Attitudes

In section 1, I introduced a tight constraint on when an individual can have given propositions as objects of their beliefs, the Stability Constraint for Belief. It is natural to wonder whether we might generalize it for the other attitudes. Ultimately, I think we can, but it will have to be very different than the constraint for belief. My aim in this section is to convince you of that. I'll start by presenting some intuitive examples, and then I'll proceed more systematically.

One day after long hours working, I walk to my car and see it banged up. On the windshield there's a note that just says "sorry!", with no name, number, or anything else that would let me get the necessary insurance information. I think to myself in the moment that I hate whoever was responsible. My state of mind could be characterized, I claim, as hating whoever was willing to damage my car without apologizing. I didn't know who this was, in the ordinary and colloquial sense of 'know who'. But I have some suspects in mind: two people, A and B, who work at the same office as me had dents in *their* cars in roughly the place you'd expect were they the culprit. So I think to myself: if it was A, I hate A; and if it was B, I hate B; and if it was someone else, I hate *them*. Suppose it was A. Then I really do, it seems, hate A. After all, it was appropriate for A themselves to think something like "I bet the car's owner really hates me".

Again, that's the material mode; to say the same more briefly in the formal mode, the following sentences capture things I might appropriately think or say:

(6) I hate whoever was willing to damage my car without apologizing.

(7) If it was A, then I hate A; and if it was B, then I hate B; and if it was someone else, then I hate them.[16]

This case is a counterexample to something we might call the Stability Constraint for Hatred. There are no externalist mechanisms for determining the content of my attitude that would allow my hatred state to be modally unstable. I could, perhaps, introduce a name for whoever did it. If I did that, which I need not have, then this case might not have been a counterexample to the Stability Constraint for Hatred; as it stands, it is, though.

This maneuver is not available when it comes to other non-doxastic attitudes. Suppose after cooling down a bit, I think to myself: if they had a good reason to be in a rush, some kind of emergency, then I hope that they don't feel bad about damaging my car; but if they didn't, then I hope they do feel at least a little bad about it. If I'm right that I can hope in this way, then my hope state would modally track a property of propositions: that they not feel bad, in worlds where they had a good reason to rush; and that they feel at least a little bad, in worlds where they didn't. There is no externalist content-determining mechanism playing any important role here that I can see. There are no names I use to baptize anything, and no use of terms like 'water' relevant to making this modal tracking possible. So it is a straightforward counterexample to what I'll call the Stability Constraint for Hope. Similar cases can be constructed for other attitudes. All of these are appropriate things to think: "If they had a good reason, I want them not to feel bad; but if they didn't, I want them to feel a little bad"; "If they had a good reason, I'm afraid that they might feel too bad; but if they didn't, I'm afraid that they won't feel bad". These are counterexamples to Stability Constraint for Desire and Fear.

[16] A referee helpfully points out that this sounds worse in the third person, e.g., 'if it was A, then Mary hates A; and if it was B, then Mary hates B.' But they don't always sound bad: 'Mary hates whoever damaged her car. So if it was A, then Mary hates A, and if it was B, then Mary hates B' sounds pretty good to me. Here's another case. Imagine Oscar knows his partner's about to give him a dog, and he knows it'll be either a poodle or a corgi, and he hopes for whichever of them is hypoallergenic, but he doesn't know which breed is hypoallergenic. It seems perfectly appropriate for someone to say 'ah, if the poodle's the hypoallergenic one, he's hoping that he gets a poodle!' And if *I* know that it's poodles that are hypoallergenic, then I can even say 'ah, then he hopes for the poodle!'. Still, why is it hard in some of the other cases? One possibility is that, as we'll see in section 6, we ought to be extra cautious with these ascriptions sometimes when we don't know what competing considerations might interfere with their truth. This is an asymmetry between self and other, because we know what possibilities we're considering but others don't. Another, complementary possibility is that we attribute attitudes to ourselves less cautiously than we attribute them to others.

Let's do things more systematically. I would like to draw a contrast with belief and these other attitudes. I'll focus on hope, though I think the points generalize. It's important to make sure the contrast emerges from exact linguistic parallels. Start with examples with belief:[17]

α. *'That'-clauses*.

(8) I believe that there was a unique individual, Homer, who wrote both the *Iliad* and the *Odyssey*.

β. *Object-denoting names*.

(9) I believe *H*.

γ. *Descriptional singular terms*.

(10) Some people say that there was a Homer, and some people say that there wasn't. I believe whichever of those claims is the true one.

δ. *Conditionals with attitude verbs in the consequent*.

(11) If there was a unique individual, Homer, who wrote both the *Iliad* and the *Odyssey*, I believe that there was.

In the relevant circumstances, none of (8), (9), (10), and (11) are assertible, and the natural explanation is that they aren't because the speaker cannot rationally have the attitudes that would make them true.

To show a contrast, I need to exhibit examples of α–δ's types with 'hope' that are felicitous in analogous circumstances. Ideally, where I can't do that, I would also be able to explain why I can't. Here's the case I'll work with. Suppose John is deliberating about whether to have children, and the thing that he most cares about is whether having them will bring him the most happiness or not. (For simplicity's sake let's assume with him that there won't be ties.) Suppose John says to himself 'one possibility is that I have children, and the other is that I don't. Let's call the possibility that'd bring me more happiness *C*'. Finally, let's assume that having kids would bring him the most happiness, but also that John doesn't know that (and knows he doesn't know that). So, with that said, here would be the parallels:

α. *'That'-clauses*.

(12) I hope that I will have children.

[17] Thanks to a referee for pushing me to be more explicit about this, and for the 'John' example.

β. *Object-denoting names*.

(13) I hope for *C*.

γ. *Descriptional singular terms*.

(14) Either having children would bring me the most happiness or not having children would. I hope for whichever of those actually would bring me the most happiness.

δ. *Conditionals with attitude verbs in the consequent*.

(15) If having children would bring me the most happiness, then I hope that I have children.

I'll start with γ and δ because α and β introduce complications.

(10), I claim, sounds infelicitous in the relevant circumstances, one in which the speaker has no relevant evidence. There are possible situations in which someone who says (10) *would* have special evidence, where they're saying the second sentence of (10) to be coy. That would be felicitous. But in the relevant circumstances, it's not. In contrast, (14) is a perfectly felicitous thing to say even in John's circumstances as specified. We wouldn't bat an eye. So we have one instance of the desired contrast between belief and non-doxastic attitudes like hope.

You might worry that (10) and (14) are not properly parallel, for two reasons. First, (10) uses 'believe', but (14) uses 'hope *for*' rather than 'hope'. But if 'hope' and 'hope for' express different underlying attitudes, that breaks the parallel. But this difference shouldn't make us think the examples aren't properly parallel. When 'hope' takes a singular term, as here, it seems that the preposition 'for' is syntactically required in English. On a view I'm attracted to, 'hope for' is the fundamental form, and when there's a 'that'-clause, the 'for' must be suppressed, whereas 'believe' is the fundamental form and thus freely takes 'that'-clauses without needing to suppress anything.[18] As for the sameness of the attitudes expressed: it seems to me that John hopes for rain iff John hopes that it rains. 'John hopes for rain, but doesn't hope that it rains' and 'John hopes that it rains, but doesn't hope for rain' sound bizarre. Perhaps with a lot of work, one can dissociate the implicit locations associated with each conjunct. But the natural, coordinate interpretations sound very bad to me. Other examples: I hope for peace, so I hope that there's peace; and I hope that there's peace, so I hope for peace. I think the syntactic evidence combined with these considerations give us

[18] This is Nebel (2019)'s view; for syntactic evidence, he appeals to Dixon (2005). For example, note that one way, perhaps the only way, to convert 'John hopes that it rains soon' into a passive sentence is 'that it rain soon is hoped for by John'. Similar things happen with 'wish [for]' *vs.* 'wish that' and 'complain [about]' *vs.* 'complain that'.

pretty good reason to think that hoping and hoping-for are the same underlying attitude, and that 'hope' and 'hope for' appear in different syntactic environments for purely syntactic reasons.[19]

The other worry is that 'believe' in (10) took 'whichever...claims...' as its argument, whereas 'hope for' in (14) doesn't take claims or propositions but some different object, maybe possibilities or states of affairs. There are great, ongoing disputes about just what the class of so-called propositional attitudes—attitudes with propositions as objects—is. For a while it was common, I think, to think the class was pretty large, encompassing belief but also, e.g., fear and hope.[20] Lately that view has come under attack, since while it sounds fine to say 'I believe the proposition that mathematics reduces to logic', it sounds bad to say 'I hope [for] the proposition that my friend gets the job she wants'.[21] For my purposes, it doesn't matter whether belief and hope take exactly the same metaphysical types of objects. I claim just two things. First, the acceptability of (14) and the unacceptability of (10) in relevantly similar circumstances suggests that hope (for) exhibits non-constant modal tracking and belief doesn't. Second, this example doesn't employ any externalist content-determining mechanisms; 'wh'-'ever' seem to essentially behave just like definites, just with ignorance presuppositions.[22]

Let's turn now to δ, the conditionals. (11) is clearly infelicitous in the relevant circumstances, whereas (15) is clearly felicitous. So, once again, we have parallel cases where 'belief' sounds bad but the analogous example with 'hope' sounds good.

Here I should note that it's actually not too hard to find conditionals with 'believe' in the consequent that sound good in the relevant circumstances. Thus, suppose I say or think of some friends:

(16) If they stole my lunch, then I think they're in big trouble!

(Maybe I'd think that because I think I'd find out eventually and call them out on it.) I can say or think this even when I don't know whether they stole my lunch. In the appendix, I argue that these conditionals have to be interpreted differently than analogous conditionals with, e.g., 'hope', in a way that sustains the disanalogy between them that I've been developing. Here I'll just skip to the results of that discussion: conditionals like (16) are acceptable

[19] To be clear, I doubt there is a general algorithm from getting from $\ulcorner S$ hopes for $o\urcorner$ to something of the form $\ulcorner S$ hopes that $p\urcorner$, where 'o' stands for some object (possibility, etc.) and 'p' for some proposition. But sometimes it's pretty clear that the hopes are the same, i.e., that there is only one underlying attitude multiply described, as with the rain and peace cases.

[20] See, e.g., Salmon and Soames (1988, 1).

[21] For a good discussion, see, e.g., Prior (1971), Moltmann (2003) (who believes even belief doesn't take propositional objects), and Merricks (2009).

[22] See, e.g., Jacobson (1995). That's the orthodoxy, anyway, but the other main alternatives, namely that they behave like universal quantifiers or like indefinites, wouldn't change the point. See Šimík (forthcoming) for a comprehensive recent discussion.

only with an interpretation as something other than a conditional with an attitude verb in the consequent. See the appendix for the semantic and pragmatic motivations for saying that.

Conditionals like (15), on the other hand, have true interpretations as 'ordinary' conditionals—they are what they appear to be. That is, they have the form 'having children would bring me the most happiness → I hope that I have children'. We can even use (14) to argue for this. Given what I said earlier about the relation between 'hope' and 'hope for', I take (14)'s logical form to be or to entail '[every x: x is either the possibility that John have children or the possibility that John not have children](x would bring John the most happiness ⊃ John hopes for x)'.[23] This logical form *entails* the ordinary conditional reading of (15), given our assumption that having children would make him happiest. For according to the most prominent tradition in the semantics for attitude verbs like 'hope', ⌜S hopes that p⌝ is true iff S prefers their p-compatible doxastic possibilities to their p-incompatible doxastic possibilities.[24] If S hopes for the possibility in which they have children, presumably they prefer the doxastic possibilities in which they have children over the ones in which they don't. (To see this, think of how odd it would be for someone to say "I hope for peace and I hope for war".) So, by centering, according to which the closest possibility to the actual possibility is the actual world, on the orthodox Stalnaker semantics for conditionals among many others, (15) follows.

Now, John can also say this:

(17) If not having children would bring me the most happiness, I hope I don't have children.

That doesn't follow yet (at least, for conditionals stronger than the material conditional). But it does if we add that John wouldn't change in hoping for *whatever* would bring him the most happiness if, unbeknownst to him, not having children would bring him the most happiness. What he hopes for just is whatever would bring him the most happiness, and

[23] I need the inference from ⌜I bear Ψ to whatever [etc.] is F⌝ to ⌜I bear Ψ to o⌝, when o is F, to be good. An analogue of that inference fails when ⌜Ψ⌝ is an attitude verb, we replace 'whatever' with 'a' or 'some', and we read the whole thing *notionally*; 'I'm looking for a horse' can be interpreted relationally ('There is a horse such that I'm looking for it') or notionally ('I'm looking for a horse, but no horse in particular'), so we can say both 'I'm looking for a horse' and 'there's no horse such that I'm looking for it'. Maybe I'm assuming relational readings, when they might better be interpreted notionally. But the inference does not seem to be blocked for 'wh'-'ever' phrases. Suppose they have the semantics of definite descriptions. 'I'm looking for the only horse in town' seems to imply I'm looking for o, where o is the only horse in town. We can then say 'ah, you're looking for Clyde!', but we can't say that in the 'a'/'some' case. Even for universal quantifiers the inference doesn't seem blocked. ('Ah, you're looking for these horses here.') Finally, the inference from 'I'm looking for the only horse in town' to 'if Clyde is the only horse in town, then I'm looking for Clyde' is itself intuitive, and it seems to require the inference I need to be good, too. So even if we read the relevant utterances notionally (⌜no F in particular⌝) (or 'in mind'), the inferences I want are not blocked.

[24] See, e.g., Heim (1992) and Anand and Hacquard (2013, 31, example 51).

this general hope of his won't go out of existence depending on which of the alternatives would in fact do that. That is, the following is intuitive in this case:

(18) If having children were the thing that would bring John the most happiness, John would still hope for whatever would bring him the most happiness.

From (14) and (18), both (15) and (17), and hence their conjunction, follow. So the ordinary conditional reading should be available and even true in the circumstances.

Now return to α and β. I cannot say (8) felicitously, nor can John (12). To that extent there's no difference between them here. But even were (12) true in the specified circumstances, it's not that hard to see why he wouldn't be able to say it felicitously: he can't know (or rationally believe, etc.) that he hopes that he has children, even if he does hope that. So by the knowledge norm of assertion or related norms,[25] he can't assert it in his circumstances.[26] So both (8) and (12) are infelicitous in the circumstances, but for different reasons.

Let's finally turn to β, i.e., (9) and (13). (9) is not felicitous in the circumstances, since it would not be felicitous to ascribe a clearly irrational belief to oneself (without some explanation, at least). (13), on the other hand, seems felicitous to me in the circumstances. If someone said it, I would interpret them similarly to how I would interpret their saying (14), given my knowledge of how '*C*"s reference was fixed. I wouldn't put much weight on this case by itself, though, because (13) is unusual for natural language. Luckily we have γ and δ to make the points, too.

So, (14) and the conjunction of (15) and (17) are both counterexamples to the Stability Constraint for Hope. Analogues are not available for belief. It's easy to construct cases like these for most non-doxastic attitudes, but I will mostly focus on hope, anger, hatred, and admiration. Put generally, for many non-doxastic attitudes Ψ, Stability Constraint for Ψ is wrong. This is puzzling in itself. One possible lesson to learn from Stability Constraint for Belief is that attitudes in general cannot modally track non-constant properties of propositions across worlds, or that they need to be determined

[25] See, e.g., Williamson (2000) and Lackey (2007).

[26] But why can't *we* say that he hopes that he'll have children? We might know that that's what would bring him the most happiness, or think we know. Yet something feels weird in saying that hopes that he does. It can feel more appropriate in some circumstances. But it can be especially hard to just stipulate that we *know* he'd be happier with children. How would we? People are different. But it can happen in other cases, e.g., with Oscar and the hypoallergenic breeds, we can say 'ah, so he hopes he gets a poodle'. Since it's *obvious* and objective which of the breeds are hypoallergenic, it's easier to make the attitude ascription. That doesn't mean they're not true in the other cases, just that we take knowing to be harder, so assertions are often infelicitous, too. So even in the John case, we can say it if we mark our uncertainty: 'John hopes for whichever possibility would bring him more happiness. I'm pretty sure having children would bring him more happiness, so I guess he hopes for children'. ('I guess', like 'must', marks uncertainty and inference, making this sound better. Consider, e.g.: 'I guess he left already' when you see his shoes missing.)

by some externalistic mechanism if they do. But these examples show that that isn't true of all, or even as far as I can tell for most, of the attitudes. There is some characteristic of these non-doxastic attitudes—perhaps their *being* non-doxastic attitudes—that makes them differ from belief in this regard. That wants explaining. But before we attempt such an explanation, we need to reckon with some significant complications in the phenomena.

Say that a property F *mediates* a person S's bearing Ψ to o just when, first, S bears Ψ to o in w *because* o is F in w, and second, for some relevant set of worlds $W = \{w_1, ..., w_n\}$ such that o_1 is F in w_1, ..., o_n is F in w_n and such that S is relevantly internally identical in w and each w_i, S bears Ψ to each o_i in each $w_i \in W$ because they are F in w_i. Thus, suppose I'm angry at whoever was willing to damage my car without apologizing; then my anger at the different people who did it in different worlds is mediated by the property of being willing to damage my car without apologizing. F-ness's mediating S's bearing Ψ to o can generate counterexamples to Stability Constraint for Ψ just when the relevant worlds are ones in which some o_i is different from o.

Given this new jargon, we can formulate a hypothesis based on the failures of the different Stability Constraints: for any attitude Ψ, if the Stability Constraint for Ψ is false, then *any* property F can, in principle, mediate Ψ-type attitudes. That is the simplest new generalization: an extreme constraint on doxastic attitudes like belief, and extreme freedom for the rest of the attitudes. Unfortunately the hypothesis does not seem to be true. Very roughly, properties can be too "thin" to make attitudes involving them possible.

Suppose I become aware of my irascibility, and looking to not want to regret hastily-formed attitudes, I come to want to only hate people I *ought* to hate, if there are any people like that. But despite *wanting* to hate in that way, there seems to be some real difficulty in actually doing so. Speaking non-hypothetically for myself now, there are some attitudes I *do* want to have only fittingly—anger, hatred, and admiration all come to mind. Yet it seems to me that I can't do it in the way that I can, say, admire whoever sacrifices a lot of time and money to help people who need it. Perhaps I can come to hate only those I ought to hate in some way like this: I meditate on when hatred is fitting, if ever, and before I ever allow myself to hate someone, I do thorough empirical investigation to see whether they satisfy the conditions. I can do it then. But notice how different that is from admiring whoever sacrifices a lot of time and money to help people who need it: I can form that sort of admiration without going through the whole process the other thing seemed to require. Speaking generally now, non-doxastic attitudes seem to need thicker properties than just $O_\Psi = \lambda x$.I ought to bear Ψ to x.[27]

This sort of property isn't the only one to present this sort of difficulty. Suppose I know

[27] $\ulcorner \lambda\alpha.\sigma \urcorner$ can be read as $\ulcorner$the (smallest) function that, given α as input, outputs *true* when condition σ is met$\urcorner$, i.e., the *property* α has such that σ. See Heim and Kratzer (1998) for more.

that my friends are united in their hatred of someone, but I'm in the dark as to who it is. It turns out that it's someone I'm unfamiliar with, let's say C, a member of a cricket team I've never heard of. He blew an important play, but I don't know that; I don't know *why* my friends hate him so much. Their hatred is a mystery. It seems to me in that case, that I cannot hate C *because* I hate whomever my friends hate, even if I think that my friends are judicious in their hatred in the sense that, almost always, if my friends all hate *X*, then *X* ought to be hated. It seems that I have to have some sense of what C *did* and maybe a sense of why it was bad in order to hate him, but in the envisioned circumstances I don't have this sense.

Look next at admiration again. As I said, I can admire whoever sacrifices a lot of time and money to help people who need it. I know this, because I more or less think I do. (Or I do with some qualifications, like that they have not behaved horribly in other respects.) But I cannot admire whomever I ought to admire, or whomever my moral hero admires in the by-now familiar relevant circumstances, i.e., without going through the process I outlined above.

Checking in with the formal mode once more, I will just say that these utterances sound bad to me without having gone through some similar process:

(19) ???I hate (/admire) whomever I ought to hate (/admire).

(20) ???I hate (/admire) whomever my friends hate (/admire).

(21) ???I admire whoever is admirable.

(22) ??If I ought to hate (/admire) D, then I hate (/admire) them.

Generally, then, the attitudes Ψ that falsify Stability Constraints for Ψ are not *un*constrained in the properties that can mediate them. It is tricky to say exactly what properties can and cannot mediate these attitudes; to say that "sufficiently thick" properties are the only ones that can is not really even to give a theory, much less a satisfying one. Rather than present a specific generalization, what I will do is argue for a theory first, and then derive a generalization from the theory.

To introduce my solution, it helps to think about choice. I'll start there, and then I will generalize the model to the other attitudes. What we will get is an alternatives- and choice-based—I will say *hairetic*—theory of all the attitudes, one that can explain the phenomena in this section. I will devote the rest of the paper to developing this theory and exhibiting its explanatory power.

3 When Can We (Rationally) Choose to ϕ?

In this section, I will present a structurally analogous phenomenon similar to the ones I started with in section 1. It will, as I said, involve choice. I will then present a principle that, I claim, explains the phenomenon. Then in sections 4 and 5 I will use this principle to explain the phenomena from section 1.

By 'choice' I mean a process by which a person selects an option from an option set, on the basis of reasons, where the rationality of the choice itself depends directly on how the option set was constituted. (I take that to be broadly in line with at least some of its ordinary meanings.) A choice is always *among options*. Suppose an agent S has three possible options: either to ϕ_1, to ϕ_2, or to ϕ_3. This might be pushing buttons, or sending a paper to one of three journals, or whatever. Let's assume that the 'ϕ_i's are the descriptions S uses to think of her options. Corresponding to these descriptions there will likely be what we can call *modes of presentations* or *guises*, but since whether these things exist and what their nature is if they do is so controversial, I will pitch my discussion at a linguistic level as long as I can get away with it. I say this because it is more or less agreed by most people that we form intentions to ϕ *under a description*,[28] but what that really amounts to is *either under a description, or a mode of presentation, ...* I will also sometimes use the generic term 'specification'. Now suppose S also knows that one of them would be strictly better than another, which is itself strictly better than the third (let's say $\phi_1 > \phi_2 > \phi_3$); S's credences are equally distributed over all the possible orderings of ϕ_1–ϕ_3. S is in a tough choice, let's suppose, and S's deliberations are made no easier if she thinks to herself: "let 'ϕ^*' name the best of the three alternatives. So I ought to ϕ^*." That is so despite the fact that, in some sense, that way of thinking about the problem *does* allow one to identify the thing to do, ϕ_1. Intuitively S cannot choose to ϕ^* "*as* ϕ^*", but then again, S can very well ϕ_1.

Apart from the restriction to three options, this is a generalization of the problems I pointed to before. That will require argumentation, though, because it will require seeing the attitudes as I do, as *choices among options*. The differences between the different attitudes will then arise from differences among the options themselves, and the way in which option sets are generated from goals that arise quasi-functionally. Seen from this perspective, the present problem is simply the problem with H and N with which I began, transposed to choice. What we really want to say is that S can't rationally choose to ϕ^* *because* ϕ^* is F^* across a relevant set of worlds. That is, S can't rationally choose to ϕ^* across modal space without somehow relevantly changing internally, but the question is *why not*.

Here is a possible answer. If S *tries* to choose ϕ^*, we should expect her in the long run to be successful only about one third of the time at ϕ_1-ing (or, across modal space, where

[28] See, e.g., Davidson (1963) and Anscombe (2000).

the different options are ordered differently, at doing whatever option is F^* in that world). Suppose, though, S has a bias in her selection mechanism; though her credences really are equally divided about just which of ϕ_1–ϕ_3 that ϕ^* is, she just tends to ϕ_1 much more than the others. Still, she should regard *herself* as only about 1/3-likely to ϕ^*. It is relatively natural to think that one cannot choose to ϕ unless one should regard oneself as sufficiently likely to *actually* ϕ.[29]

There might be something right to that explanation, but it has to be pursued carefully. After talking to some oracle or demon, S might have reason to think she will ϕ^* in the actual world, and will do whatever is F^* across a wide range of worlds. It's not enough for S's credences to be such that $Cr_S(<\phi_1 = \phi^*>|<S\,\phi_1\text{-s}>)$ is very high. Rather, to rationally deliberate with ϕ^* as an option, and to choose it, she needs to know *that* ϕ_1 is ϕ^*. Since she can't rule out ϕ_2 or ϕ_3 as ϕ^*, she can't know that ϕ_1 is ϕ^*, and thus could not rationally deliberate with it or choose it.

Distinguishing x and y seems to be relative to how x and y are described or, more generally, presented. Supposing that distinguishing x and y is the same as knowing that x and y are different, that means distinguishing is relative to descriptions or presentations.[30] S can know that ϕ^* and ϕ_2 are distinct under the description of ϕ_2 as the (actual) second-best of S's options, for example. Yet that doesn't actually make it so S can choose ϕ^*, since, well, S doesn't know that ϕ_1 isn't the second-best of her options.

Here's a way out of this difficulty. Suppose S's option set is $\Phi = \{\phi_1, ..., \phi_n\}$. And suppose the ϕ_is have descriptions or modes of presentations *in terms of which* S will translate her choice into action. Thus, ϕ_1's action-relevant description might be something like "give \$1,000 to X charity at t". More generally, I will call the descriptions or modes of presentation of S's options under which S will translate her choice into action the *active* descriptions and modes of presentation of S's options. Now, S might *also* think of ϕ_1 under the description "the best option in Φ". But S would be irrational to deliberate about or decide which option in Φ to choose with descriptions or modes of presentation like that if she doesn't know which active description or mode of presentation that description corresponds to. Thus, it is irrational for S to deliberate about whether to give \$1,000 to X charity at t with the description "the best option in Φ".

It'll be helpful to put this in terms of a general principle. S's deliberation or decision is *presentationally clear* just in case, for each way S thinks about her options—be they descriptions, modes of presentation, etc.—S is able to match that way of thinking with the *active* way of thinking (description, mode of presentation, etc.) that corresponds to the same action. Then:

[29] This view traces back to Grice (1971) and Harman (1976). For influential criticism, see, among others, Bratman (2009).

[30] See Williamson (2013, chapter 1) for a discussion of this view.

CLARITY. In deliberating and deciding about which option in $\Phi = \{\phi_1, ..., \phi_n\}$ to choose, S's deliberations and decision are rationally required to be presentationally clear.

Before I explain why I think CLARITY is right, I want to clarify what it doesn't say. It doesn't, for example, require agents to know, of any property F that would make a given action ϕ more or less good (etc.) if ϕ were F, that ϕ is F (or that it is not F). It also doesn't require agents to know that the Evening Star is the Morning Star.[31] CLARITY only forbids using descriptions with $\ulcorner F \urcorner$ or corresponding modes of presentation—specifications—in deliberating about or deciding what option to choose when there are active specifications that don't use $\ulcorner F \urcorner$ in that way and where it is unclear to the deliberating agent to which active descriptions the ones that use $\ulcorner F \urcorner$ that way correspond.

So, why should CLARITY be true? The *function* of choice is to link deliberation and decision to behavior. The problem isn't that, when the choice situation is not clear to an agent, the agent is *less likely* to do the action they chose to do. It is, rather, that their deliberation and decision are less *connected* to what they ultimately do. Deliberation makes us choose the best and, at least in principle, makes what we go on to do *intelligible*. When a deliberation and decision are unclear to S, they are unable to accomplish that task.[32] When an agent thinks it's very likely that d_1 and d_2 both correspond to option ϕ without knowing it, either d_1 or d_2 (or possibly both) should be thought of as standing for potential *properties* of an option; such attributions of properties to objects should be handled in the normal way, with credences over propositions attributing the properties to the different options.

CLARITY rules out the case of the choice between ϕ_1–ϕ_3 as stipulated: S uses a description with 'ϕ^*' or corresponding mode of presentation while also using descriptions with 'ϕ_1'–'ϕ_3' or corresponding modes of presentation. CLARITY requires S—if she wishes to use descriptions involving 'ϕ^*' or corresponding modes of presentation—to *rule out* similar descriptions with 'ϕ_2' and 'ϕ_3' or corresponding modes of presentation as corresponding to the option 'ϕ^*' is a description or mode of presentation of. Of course, S could satisfy CLARITY by ensuring that descriptions with 'ϕ_1'–'ϕ_3' or corresponding modes of presentation are not used in her choice situation. Does that mean that, in such situations, she *can* make a choice using descriptions or corresponding modes of presentation with 'F^*'? It depends on what *exactly* her different options are, and this she might not be in a position to choose.

[31] We need to individuate actions in a relatively fine-grained way, so that going to Hesperus and going to Phosphorus can count as two different options for an agent that doesn't realize they would involve the same physical behaviors, but not so fine-grained that we can't have distinct modes of presentation of the same option. But this is a natural way to think of options, I think. Thank you to a referee for making me think about this.

[32] For a philosopher friendly to this thought, see Korsgaard (2009), among many others. Arpaly and Schroeder (2014, chapter 2) deny that deliberation has a *special* role in making action for reasons possible, but wouldn't deny, I think, the kind of minimal functional role I ascribe to deliberation.

In this section, I presented a phenomenon similar in structure to the phenomenon in section 1, and have used Clarity to explain it. In the rest of the paper I will assimilate the other cases, of belief and of the non-doxastic attitudes, to the choice case. I will argue that their particular behavior arises from the kinds of choices they are, i.e., from the kinds of options they are choices among.

4 A Hairetic Theory of Doxastic Attitude Formation

In this section, I will argue that doxastic attitudes, in particular belief, ought to be seen as choices among options.[33] To do so, I will appeal to some previous work on knowledge and belief. In the next section, I'll apply the hairetic theory of belief formation, along with Clarity, to resolve the issues from section 1.

Since at least the seventies, many epistemologists have thought that whether it's true that S knows that p depends, somehow, on S's relation to some special *subset* of the set of $\neg p$ possibilities. There are intense ongoing debates among those epistemologists, and I want to sidestep those controversies as much as possible. Here I will reproduce some of their motivations, and argue that a hairetic theory of belief formation can explain them naturally and easily. It is not the only theory that can do so; perhaps some variety of relevant-alternatives contextualism can, too. According to that sort of theory, $\ulcorner S$ knows that $p\urcorner$ means, roughly, that S can "rule out" (in a sense to be specified) some relevant subset of $\neg p$ possibilities. Strictly speaking the theory I propose will not be in direct conflict with that sort of contextualism, though my theory might make it less explanatorily necessary.[34]

What, then, are those motivations? Here are three prominent ones:

- *Fallibilism*. As I said, some philosophers have wanted it to be the case that knowing that p does not require ruling out every alternative incompatible with p. Thus they want it to be possible that I can know that I have hands even if I can't rule out that I'm a handless BIV for some appropriate sense of 'rule out'. Call that minimal position *fallibilism*. A popular way of being a fallibilist is to think that only some p-incompatible alternatives are *relevant*. This tradition, the relevant alternatives tradition, then subdivides along a number of different dimensions.[35] What matters for

[33] Most theories of belief formation aren't framed this way. But there are exceptions. See, e.g., De Sousa (1971, 59), who describes beliefs as *bets on the truth*. But he does not consider the importance of options. The epistemic decision theorists from footnote 1 likely also presuppose it.

[34] The main job I see left for relevant-alternatives contextualism in light of what I say here is to explain ascriber effects, the intuitive truth conditions of knowledge attributions that vary according to the ascriber's (or evaluator's) context rather than the actual circumstances of belief formation. I leave that work to others (see Lewis in the footnote below, e.g.).

[35] Here are some of those dimensions. I mentioned contextualism before; so, does the relation of relevancy differ according to the ascriber (or evaluator)? According to some, e.g., Lewis (1996), it does. Most relevantly

me is this sensitivity to relevant alternatives, since I take my own eventual "option sensitivity" account to capture it, including for belief.

- *Question Sensitivity*. Whether I know that *p* seems sensitive sometimes to what some relevant question is. To take an example from Schaffer (2007), I can know that the person on TV is George W. Bush when the background question is whether the person on the TV is Bush or Beyoncé, but not if the question is whether it's Bush or famous Bush impersonator Will Ferrell.[36]

- *Linguistic Motivations*. 'Knows' seems to associate with focus in the way that expressions like 'only' and 'always' do. *e* associates with focus roughly when *e* introduces truth-conditionally-relevant *alternatives*. E.g., 'I only went to the BANK today' introduces alternatives like that I went to the grocery store, mall, etc., and the sentence as a whole requires that these all be false.[37] Thus:

  (23) a. I know that Clyde SOLD his typewriter to Alex.
  b. I know that Clyde sold his typewriter TO ALEX.

  Roughly, (23a) can be true in a situation where the question is "what did Clyde do with his typewriter?", but not when the question is "to whom did Clyde sell his typewriter?" (and *vice versa*).[38]

There might be others, but for now these are the ones I will focus on.

I claim that the same motivations apply to rational belief. (Indeed, we can even think of the alternatives sensitivity of knowledge as arising from the alternatives sensitivity of rational belief, at least if knowing that *p* entails rationally believing that *p*. I won't pursue that in detail in this paper.) Start with fallibilism, and return to the BIV example. I believe that I have hands, and I think that belief is rational. But—depending on the right

for our purposes, for him an alternative is relevant if the ascribers are attending to it. Other dimensions include what "ruling out" comes to and whether the theory secure epistemic closure. Given phenomenal indistinguishability account of ruling out and his contextualism, he accepts closure, but only all within the same context; when we change what we ascribers are attending to, we'll be able to say first, ⌜S knows they have hands⌝, and second (when we change what we're attending to) ⌜*S* doesn't know they're a handless BIV⌝. Dretske does not think relevancy is context-sensitive in this way, and given his conception of relevancy and ruling out, he'll deny closure *within* a context. Stine (as I read her) is a contextualist like Lewis, but she denies that mere attention makes an alternative relevant, and thus will countenance fewer cross-contextual apparent failures of closure than Lewis will. See Holliday (2015) for an excellent and thorough discussion of fallibilism and relevant alternatives theory, including a grounding of the different dimensions here in some even more fundamental dimensions.

[36] See also the examples and discussion in Schaffer and Szabó (2014).

[37] See, e.g., Rooth (1992).

[38] For the examples, see Dretske (1981, 373). See also Schaffer and Szabó (2014) for similar examples and for more linguistic motivations that I spare the reader for space.

conception of evidence and of ruling out—I cannot rule out my being a handless BIV.[39] A relevant-alternatives theorist will allow us to have knowledge despite not having ruled out incompatible possibilities that we are properly ignoring, so a relevant-alternatives theorist should allow us to have rational belief despite not having ruled out those same possibilities. Otherwise, they will be committed to thinking that we can know p without rationally believing p.[40]

Question sensitivity works similarly. I can rationally believe that the person on the TV is George W. Bush when the question is whether it's him or Beyoncé, but fail to rationally believe it when the question is whether it is Bush or Ferrell. Once again, to deny this would be to say that we can know that it's Bush in the first case without being able to rationally believe it. That is unattractive.

Finally, the (other) linguistic motivations apply to belief, too. In particular 'believes' associates with focus:[41]

(24) a. I believe that Clyde SOLD his typewriter to Alex.

b. I believe that Clyde sold his typewriter TO ALEX.

Someone who is in a position to say (24a) need not be in a position to say (24b): (24a) requires the agent to have "ruled out" non-selling possibilities (e.g., loaning, etc.), whereas (24b) requires the agent to have "ruled out" non-*Alex* possibilities, (e.g., to Bert, etc.).

A hairetic theory of belief *formation* is natural in the light of these theories. It is impossible to say whether the choice to ϕ is rational without knowing what the overall option set Φ was out of which the choice to ϕ was made. A choice to ϕ, say to buy a particular plane ticket, can be rational given one option set (say, among relatively expensive tickets), but not in another (say, among relatively cheap tickets). This is just the relativity of rational choice to options. All of the examples I gave can be understood this way, too. For example, if the choice of what to believe about who the person on TV is is between the proposition that it's George W. Bush and that it's Beyoncé, if anything, then it can be rational pick the proposition that it's Bush; but if it's between that it's George W. Bush and that it's Ferrell, if anything, then it is not rational to pick that it's Bush. The explanations of the other data are fairly routine given this template. The argument for the naturalness of the hairetic theory of belief formation, then, is just that it can elegantly explain all this data about belief and even, possibly, knowledge, though again the latter is not my aim here. If our option set

[39] For those who think with Williamson (2000) that our evidence is our knowledge, this won't be plausible. But such people aren't fallibilists, for that very reason.

[40] Lewis himself was a weird case, in that he denied the link between knowledge and belief, and even knowledge and justification. I assume most relevant-alternatives theorists will not wish to follow him in that.

[41] See Beaver and Clark (2008, 51) for an argument that such data can be explained pragmatically; it appeals to the mechanics of question sensitivity, as will my own ultimate explanation. This once again undercuts the motivations for contextualism somewhat.

about what to believe properly excludes the proposition that I am a handless BIV, then it is rational to believe I have hands. It is likely very difficult to give a theory of when such options can be properly excluded; my point is that sometimes they clearly can be, and that makes a difference to the rationality of the belief formed in those circumstances.

So, if belief formation is a choice among options, what are the options? I will assume that they are determined by questions relevant at the time of the belief formation, with some other possibilities properly excluded in addition. On an orthodox view, question-denotations are sets of propositions; if the question is Q, then its denotation $[\![Q]\!] = \{p\colon p$ is a possible answer to $Q\}$, where a possible answer to Q is an answer Q *does* have in some possible world.[42,43] This is not the view philosophers always express, at least when it comes to knowledge: many of them have thought that alternatives should rather be conceived as *possible worlds*.[44] Nevertheless, I think the arguments for making belief relative to alternatives support thinking those alternatives come from questions. And it should be said that even among some of those who have thought of alternatives as possible worlds, they have been sometimes thought of as *coarse-grained* worlds[45]—which we can think of as actually propositions, reducing again to our question-based approach.[46] Finally, since some propositions are singleton sets of worlds, the propositional alternatives I imagine will perhaps sometimes include single-world alternatives. So I will assume going forward that questions do indeed generate the alternatives in the option set (e.g., believing, disbelieving, etc. p, which might answer Q) though with the understood caveat that forming no belief is at least typically an option, too.

So, on my conception, a person's doxastic options are coming to believe, disbelieve, adopt x credence in, suspend about, etc. the various answers to the questions she asks herself that are her relevant alternatives. Given that, I am now in a position to present the hairetic theory of belief formation:[47]

> Hairetic Theory of Belief Formation. *S forms a belief* that p only if there is a question Q such that $p \in [\![Q]\!]$ and S chooses to believe p in answer to Q rather than each $p_i \in [\![Q]\!]$ such that $p_i \neq p$.

[42] '$[\![e]\!]$' stands for the denotation of the expression *e*.

[43] See Hamblin (1958). There are other views of question-denotations in linguistic semantics and recent philosophy, but they will not meaningfully change what I say here.

[44] See Holliday (2015) for a good discussion of this point.

[45] See Lewis (1996).

[46] Yalcin (2018) makes a similar point on this score. The coarse-grained conception itself is also present in Stalnaker's work, e.g., Stalnaker (1984), though in a different dialectical context. See also, e.g., Edgington (1995, 266) and Swanson (2012, 1549–1551), among others.

[47] I am putting things in terms of belief *formation*. I could have instead presented a contextualist theory of 'belief', or a theory that was disjunctive between them. I decided not to, because I think a contextualist theory of rational choice that my account would need to rely on isn't very natural. But a contextualist version of what I do here might very well be worth pursuing!

Every belief needs to be formed, i.e., it needs to come into existence somehow. Sometimes this is the result of a conscious activity, inquiring. That activity brings us from one kind of attitude—curiosity, wondering, etc.—to a settled state of belief (or we lose our curiosity, or confidence the question has an answer we can discover). Those attitudes themselves seem to have as their object a question, the question the belief they come to at the end is meant to answer.[48] By the end, in other words, we have chosen one of the possible answers as the proposition we believe, the one that is the actually true answer to the question we started out curious, etc., about.[49]

Not every belief we have is formed so explicitly or consciously, and *a fortiori* as the result of some conscious activity of inquiring. That shouldn't worry us any more than the automaticity of many *actions* does. Sometimes a choice is obvious, and sometimes it happens sub-personally. I come to believe a friend of mine is wearing a blue shirt by looking at him and his shirt; it would be strange to describe the process leading to that belief as an inquiry, even if I formed that belief instead of the belief that it is green, red, or whatever. We should not let the quickness of the formation of the belief, or the ease of the issue, confuse us about whether it was a choice among options or not. If you ask me how many fingers you're holding up (holding some up as you ask), I won't have to inquire, and I'll say the right answer more or less immediately. But it was a question with different possible answers that prompted my choice, my answer. That said, 'chooses' certainly has voluntarist overtones, and I do not mean to suggest that belief formation is like raising one's arm in that respect. As I said in section 3, by "choice" I mean a process by which a person selects an option from an option set, on the basis of reasons, where the rationality of the choice itself depends directly on how the option set was constituted. This is a very broad notion of choice that requires no kind of voluntarism.[50]

Finally, there's the question of what *determines* the relevant question or questions. Many of the beliefs we form happen in conversation, and it is natural when they do to think that the relevant questions are what Roberts (2012) calls "questions under discussion", the (hierarchically-ordered) questions participants in the conversation are trying to answer at a given time. But some inquiries don't happen in conversation, and in such cases we will have

[48] For this picture, see especially Friedman (2013).

[49] My remarks in this paragraph might remind the reader of what Boyle (2011) calls the 'process theory' of belief: "[d]eliberation about whether P is a process that culminates, if things go well, in a judgment on the truth of P. Judgment is an occurrent act by which a subject installs a new belief in herself, or modifies one she already holds. Belief itself is not an act but a state" (5). Boyle attributes this view to, among others, Peacocke (1998) and Shah and Velleman (2005), and offers trenchant criticisms. But what I say should be neutral between Boyle's alternative, Aristotelian model and the process theory he rejects, since what matters for me is just that it is a kind of choice among options.

[50] 'Selects' has more of a connotation of 'sorting' than 'chooses', I think, as in 'such-and-such evolutionary process selects for blah'. For a good quasi-voluntaristic explication of doxastic assent, especially as connected to the Stoics, see Wright (2014).

to say that the relevant questions are ones the agent is attempting to answer in forming her beliefs, and these questions generate her options through the alternative possible answers they present.

Hairetic Theory of Belief Formation is natural, and it seems to me to be the best explanation of the data I exhibited. The full argument for it will, as I said, be that and its ability to solve our puzzle. So, to solve the problem of modally unstable belief, we need to look at one more thing: when something is a *good answer* to a question. That is what I will do in the next section. Then with Clarity, we will have the resources to explain the phenomena from section 1.

5 Good Answers and Modally Unstable Beliefs

In this section, I will argue that constraints on what counts as a good answer to a question will join with Clarity to explain the doxastic phenomena from section 1.

A question not only introduces the alternatives which determine the options among which a potential believer chooses, but it also introduces the proper *specification*—description, mode of presentation, whatever—of those alternatives. Start with a simple example:

(25) a. Did you go to the party yesterday?

b. {<I went to the party yesterday>, <I did not go to the party yesterday>}[51]

Not only is it the case that $[\![(25a)]\!] = (25b)$, but (25b) *gives the proper specification of the answer*.[52] The way a question is specified systematically determines the proper specifications the possible answers may take. As a polar question, (25a)'s two alternatives correspond to the question *rephrased* as an answer, either with or without a negation. Now consider:

(26) a. Of X_1, ..., X_n, who went to the party yesterday?

b. $\{p\colon p = \lambda x.x$ went to the party yesterday$(X_1) \wedge ... \wedge \lambda x.x$ went to the party yesterday$(X_n) \wedge \neg(\exists X)(X \neq X_1, ..., X_n)(\lambda x.x$ went to the party yesterday$(X))\}$

Depending on how exactly we should think the semantics of 'wh'-questions like (26a) should work, (b) not only gives its denotation, but also gives the proper form an answer

[51] I use $\ulcorner <\sigma> \urcorner$ for the proposition that σ.

[52] There is precedent in thinking that ways of specifying a linguistic item's denotation is thereby to give its sense—since specifications of a denotation just *are* senses. See, e.g., interpreters of Frege (and likely Frege himself, I think), such as Dummett (1973) and Kripke (2011).

might take.[53] The important point is that an answer to (26a) will have something like the form specified in (b).

In general, specifications of the question give specifications of the options under which an agent deliberates and decides. Sometimes it won't be obvious what a question-denotation is that also gives the correct specification. For example:

(27) Why is the sky blue?

I would *guess* that $[\![\ulcorner$why $p?\urcorner]\!]$ is something like $\{q\colon q = \langle p$ because $r\rangle\}$, in which case $[\![(27)]\!] = \{q\colon q = \langle$the sky is blue because $r\rangle\}$. To be sure, the general project of showing *how* the specifications of questions determines the specification of possible answers might be very difficult. But like the rest of semantics, it seems to be tacit knowledge the theorist aims to uncover and describe explicitly. In other words, we *can* do it, even if it is difficult to say exactly what we are doing.

Questions establish the specifications under which an answer takes its proper form because, given the Hairetic Theory of Belief Formation, they *frame the choice*. A question that a person is curious about presents its possible answers as answers *to it*, *as it* specifies the form a proper answer is to take. In the terminology of section 3, a question makes certain specifications of the propositions that answer it *active*. Someone who wishes to answer (25a), that is, to come to a belief that answers (25a) in particular, can pick one or the other of the set in (b), *as presented by the question*. Questions are like the descriptions of a choice that an agent is given; it is possible to conceive of the choice differently, of course, but not when one is still attempting to answer the question that framed the initial choice—unless one can connect the different conceptions of the choice, that is, can say which options (and the alternatives that determine them) correspond when conceived in the new way to the options as the question itself presented them.

With all this in mind, we are at last ready to explain why belief cannot be (rationally) modally unstable, as seen in section 1. So, let's return to H. The aim is to explain why individuals can't rationally believe H in the evidential circumstances I stipulated there. What is the question to which H might be the answer? The easiest question to check first is:

(28) Was there a unique individual, Homer, who wrote the *Iliad* and the *Odyssey*?

This is a polar question, just like (25) was. So it seems to have the denotation {<there was a unique individual, Homer, who wrote the *Iliad* and the *Odyssey*>, <there was not a

[53] According to (b), a proposition answers (35a) when the proposition, roughly, says of every person either that they do not do not fall in the extension of $\lambda x.x$ went to the party yesterday. This is called a *strongly exhaustive* reading. Other denotations are possible. See, e.g., Theiler et al. (forthcoming) for an excellent, recent survey of the possible question denotations, and a compelling theory of their own.

unique individual, Homer, who wrote the *Iliad* and the *Odyssey*>}. Moreover, an individual choosing which of either of those to believe must do so with the way I just specified those propositions (or something like that way). Supposing again that the second element is the true one, that is the same proposition as *H*. But if they tried to choose *H* as presented in section 1, the choice would not be presentationally clear, violating Clarity. After all, they don't know which of the two elements of ⟦(28)⟧ *H* is. In other words, such an agent cannot rationally come to believe *H* as an answer to (28), since the deliberation and decision leading to forming that belief would be irrational. This is exactly the result we wanted.

(28) isn't the only question to which *H* is an answer, of course. But what are the other possible questions? "Is *H* (so defined) true?" has an easy answer, that it is, but we have already seen that believing *H* is not the same thing as believing that *H* is true. Other questions to which *H* is the answer will likely all have the same problem that believing *H* as an answer to (28) has.

Speaking more generally, beliefs are formed as answers to (possibly implicit) questions. The beliefs themselves, to be rational, have to be good answers to those questions. But there is no question to which simply "*H*" is the answer. There's a grammatical reflex of this: "*H*?" is not a grammatically well-formed question. "Is *H* true?" is, of course, and as we've seen in section 1 and here, we *can* form a belief relative to that question. The Hairetic Theory of Belief Formation captures this explanatory desideratum very nicely; it is very hard to understand otherwise just *why* we can believe that *H* is true without believing *H*. So, *if* rational belief formation is a response to questions, and if questions present the specifications under which a question is to be answered, we can explain why *H* cannot be believed in the given circumstances: anything that actually answers a given question will not have the bare form '*H*'; but even if *H is* one of the answers to that question (presented a different way), Clarity rationally excludes choosing it as one's answer under that specification. To do so, agents would have to know which answer it corresponds to. In that case, there would be no Clarity violation; but at that point it's useless to think of *H* in that way, anyhow.[54]

[54] A referee mentions a very interesting possibility, with which I will have to be too brisk here. Let 'gurge gurges' be synonymous with (i.e., have the same semantic value as) the true element of {'there was a unique person who wrote the *Iliad* and the *Odyssey*', 'there was not a unique person who wrote the *Iliad* and the *Odyssey*'}. If that's possible, 'Was there a unique person who wrote the *Iliad* and the *Odyssey*?' *would* have an answer I could give in my present evidential state: I could just answer (in thought or speech) 'gurge gurges'. Thus, the reality of sentential stipulation would sit badly with my argument. First, when we know what 'does gurge gurge?' is asking because we can build it up compositionally, we know that 'gurge gurges' is *not* a good answer to the question we were asking, which we can know because no one who had the question would be satisfied with it as an answer. For a person to wonder whether gurge gurges just is to wonder whether there was a unique individual, Homer, who..., just like a person who wonders whether Julius pegasizes would be wondering—and take themselves to wonder—whether Julius was a pegasus, where by 'pegasizes' we mean the property $\lambda x.x$ is Pegasus (taking inspiration from Quine (1948)). And if 'gurge gurges' *isn't* built up compositionally, then it would not be possible for a person to ask themselves a question with it beyond 'is it true that

We can be briefer with the other phenomena from section 1. Next return to *N*. The natural question here is:

(29) Were neanderthals on average smarter than humans?

Again, this is a polar question. $[\![(29)]\!]$ = {<neanderthals were on average smarter than humans>, <neanderthals were not on average smarter than humans>}. But because the agent doesn't know which of those, so specified, is the proposition believed by the majority of anthropologists on the issue, were '*N*' to be the active specification, the choice would not be presentationally clear, violating Clarity. So an individual cannot rationally believe *N* as the answer to (29), for the same reason as before. Nor is there any *other* question *Q* that it seems individuals *can* believe *N* (so specified) as an answer to *Q*. Again, this is a good result.

Finally, consider *J*, the proposition that Julius was clever. Another polar question is natural here:[55]

(30) Was Julius clever?

$[\![(30)]\!]$ = {<Julius was clever>, <Julius wasn't clever>}. Believing *J is* possible for me in these circumstances, since the choice is presentationally clear, even if the proposition changes in different worlds depending on who invented the zipper there. This kind of doxastic modal tracking is possible—'Julius was clever' is a good answer to that question. That's because the very question is *framed in terms of it*. That is, when the question itself has elements whose content is supplied by externalist content-determining mechanisms, then good answers to them may use such mechanisms. There is no violation of Clarity here because there are no other modes of presentation of these options *other* than the 'Julius' modes of presentation. The options allow me to form the belief.

The reason agents cannot use externalist content-determining mechanisms to believe *H* is that, as defined, there *are no* externalist content-determining mechanisms in the questions to which *H* is a good answer, with its mode of presentation. As I said, there is the question $[\![Q]\!]$ = {<*H* is true>, <*H* is not true>}, which, in addition to not being a question to which *H* itself (rather than that *H* is true) is an answer, is a trivial choice and so uninteresting. This completes the explanation of the data in section 1.

gurge gurges?', which would allow one to believe that it's true that gurge gurges but not that gurge gurges, just like with *H* itself. (Nor can we use 'gurge gurges?', with upward intonation on the second word. We can't assume we can know *where* we should put the intonation to make it, syntactically, a question. If it's unanalyzable, there's no answer to that question.) Note, by the way, that I doubt we can use sentential stipulation to *hope* that plurge plurges, for some suitable sentential stipulation for 'plurge plurges'. The mechanism I will use to explain non-doxastic non-constant modal tracking is different from this one.

[55] There are other possible questions, e.g., 'who are the clever people?'. But these would work similarly to (30).

Clarity—a general constraint on possible choices—combines with Hairetic Theory of Belief Formation to solve the problem of impossible modally unstable belief. This is a considerable point in each one's favor, though I also gave independent motivations. To explain the data with the non-doxastic attitudes from section 2, however, I will need to come up with analogues to the Hairetic Theory of Belief Formation, but they must be different enough that combining with Clarity doesn't rule out hating whoever was willing to damage my car without apologizing, for example.

6 A Hairetic Theory of Non-Doxastic Attitude Formation

In this section, I will use the work from sections 3 and 4 to explain the odd phenomena from section 2 involving the non-doxastic attitudes like hope, hatred, anger, fear and admiration. (Remember that, though I think what I do here extends readily to most other non-doxastic attitudes so that I sometimes say 'non-doxastic attitudes', I only wish to make claims here about the specific ones I investigate, namely the ones I just mentioned.) To do that, I will argue that, like with belief, whether a given non-doxastic attitude was formed rationally depends not just on the object of the attitude, but also on what the agent's overall option set was.

In section 4, I gave reasons for thinking that knowledge was somehow relative to alternatives, and then showed that they more or less applied in the same way to belief. There is likely no state that stands to, say, admiration, as knowledge does to belief. But that doesn't mean I think the arguments are weaker. I'll focus on fallibilism and question sensitivity, starting with the former.[56] Consider (6) again:

(6) I hate whoever was willing to damage my car without apologizing.

I can *say* that, but suppose it was my wonderful mother whom I love dearly—surely I wouldn't hate her. Thus:

(31) I don't hate whoever damaged my car and ran away without providing information if it was my mother.

(6) should be able to be true even if (31) isn't. These kinds of examples are easy to multiply:

(32) a. I admire whoever gives a large amount of their time to helping the homeless.

b. I don't admire someone who gives a large amount of their time to helping the homeless if they only did it to impress a crush.

[56] Admiration, etc., *do* associate with focus. But the examples are less interesting or telling in this case, and so I am not going to go through them here.

The analogue of fallibilism for these attitudes is that an agent can say apparently reasonably that they bear Ψ to whomever is *F*, and the same agent could apparently reasonably deny that they bear Ψ to people who are both *F* and *G*.

The reason is that the possibility that in (6), for example, the speaker is just not entertaining the possibility that it was their *mother* who damaged their car. Similarly, in (32), the speaker is just not entertaining the possibility that people do that to impress crushes. Insofar as restricting the choice of whom to hate or admire is reasonable in this way, and it certainly seems to be, then such speakers will sound reasonable. This is very similar to our character before who believed that the person on TV is Bush because she (properly or reasonably) doesn't take seriously the possibility that it's Will Ferrell.

Next, turn to question sensitivity. It might seem that this features of alternatives-based theories ought not to exist for these non-doxastic attitudes because they are not the answers to questions. But this depends on a too-narrow sense of what question sensitivity really was when it came to belief formation. Really what I've been arguing for is that the rationality of attitude formation is sensitive to *options*. With belief, the questions we ask ourselves generate our options, because those questions present the alternatives that we might believe, disbelieve, and so on. So it is helpful to talk about question sensitivity there. While it can be helpful sometimes to think about forming admiration for someone as answering a question, e.g., 'whom should I admire?', it's mostly only a *façon de parler*. The more general phenomenon is option sensitivity. Still, questions can bring out option sensitivity, since they make certain possibilities salient and can thus shape a person's option set.[57]

So, where do we see option sensitivity? Consider:

(33) a. Do you admire giving money to a museum, rather than keeping it?

b. Do you admire giving money to a museum, rather than to a more effective charity?

One can imagine someone answering 'yes' to (a) and 'no' to (b). Of course, one way to interpret this example is as asking, first, "do you admire giving money to a museum *more than* you admire someone's keeping their money?" and, second, "do you admire giving money to a museum *more than* you admire giving to a more effective charity?". But that's not the only interpretation, or even the most natural. More natural, I think, is that what (a) asks is, roughly, "*assuming* that the person would otherwise keep their money, do you admire their giving it to a museum?" And that answer can be different from the answer

[57] I am here implicitly appealing to the fact that, because we're limited creatures, we can't be expected to think of every possibility. Our options are the things that, roughly, we should take seriously as things to do; they are tied to the subjective 'ought'. (See, e.g., Hedden (2012)). Mentioning possibilities makes them much easier to think of, then, and to take seriously. So once mentioned, often we should take them seriously, and so those possibilities often give rise to different options.

to (b) interpreted in that way, namely "*assuming* that the person might otherwise give it to a more effective charity, do you admire their giving their money to a museum?" So interpreted, this example exhibits option sensitivity; the question encourages the answerer to ignore possibilities outside of the ones it presents. This is not very different from the Beyoncé/Ferrell example. But these questions are less directly related to the attitudes formed in response to them as with belief; *belief* answers questions, whereas admiration follows beliefs about whom to admire, among other things. What matters for me is that they both exhibit option sensitivity.

So, I think an alternatives- and option-based theory of rational non-doxastic attitude formation is natural. The best explanation of this naturalness, I claim, is once again just the relativity of the rationality of choice to the options out of which one chooses. When my options are {hating whoever was willing to damage my car without apologizing, not hating whoever was willing to damage my car without apologizing, ...}, partly because it is not a salient possibility to me that my mother is one such person, then it is rational for me to pick the former. But when my options are {hating whoever was willing to damage my car without apologizing and is my mother, hating whoever was willing to damage my car without apologizing and is not my mother, ...}, the second option will be superior to the former. This perspective helps to explain the other data, too. Thus it motivates a choice-based theory of the rational formation of these attitudes.

Next we should ask what the options *are* in admiration, etc. option sets. A question-denotation is a set of (mutually exclusive) propositions—the alternatives the option set comes from. But many non-doxastic attitudes take other objects than propositions, or other objects in addition to propositions. Hatred, for example, takes both. So, what are the alternatives? For those attitudes that take both, the option sets will have subsets like {hating that cat, hating that snake, ...} and subsets like {hating that there's a war in such-and-such country, hating that X candidate will win Y election, ...}. And of course perhaps in given contexts option sets will have both kinds of objects of hatred.[58]

Recognizing this difference alone doesn't explain the difference between belief and the non-doxastic attitudes that I argued for in section 2, namely that non-doxastic attitudes allow for certain kinds of modal tracking. To explain this difference, we need to look more closely at what non-doxastic option sets are like. I will focus on admiration. Note that the question "whom should I admire?", like "what should I believe?", takes at least two sorts of answers. Consider "what should I believe about whether Trump will win in 2020?". One kind of answer would be a list of propositions, connecting his chances with the economy, e.g. Another kind of answer would tell me the *kind* of thing to believe, for example things

[58] There's an ongoing debate about whether all these attitudes can be propositional at root or not. See, e.g., Montague (2007) and Forbes (2000).

said about it by political scientists. This distinction applies to admiration, too. Consider:

(34) —Man, the Beatles were an amazing group! Whom should I admire most musically in the group?

—Paul McCartney, no doubt.

The second speaker in (34) answers the question, though perhaps they could have done so more helpfully if they had said why. A different kind of answer will give *features* of objects (individuals, likely) that make them worthy of admiration: those who give large amounts of time to help those who need it, those who achieve a lot in the face of adversity, whatever. I might not know which people *have* those features, though. This is all to say the following dialogue makes sense, too:

(35) —What should I admire in a musician?

—Taking lessons from the past while not being chained to it, superlative technical ability, and sensitivity to the broader cultural and artistic contexts in which they work.

The second speaker in (35) answers the question, though perhaps they could have done so more helpfully if they had given examples of such individuals.

It's not just that these questions have these different answers. Consider (34) again. The options associated with this question are: {admiring X most: X is or was a Beatle}. Now consider (35). The options are {admiring F-ness: F-ness is a property musicians can have}. So it's possible to admire *properties*.[59] Consider, e.g.:

(36) I admire honesty in a musician.[60]

This is different from belief, since we cannot believe properties![61] The following sound horrible, for example:

(37) #I believe truth in a proposition.

[59] I won't distinguish here between properties and tropes, or similar kinds of metaphysical entities. The important point is to get a contrast between the doxastic and non-doxastic attitudes in this regard, not necessarily to investigate the exact nature of that contrast. So consider my proposal disjunctive between what the different kinds of entities that I call "properties" might be.

[60] You may think we're really, somehow, admiring the *person themselves*. But notice: 'my favorite feature of McCartney's is his honesty; I really admire it.' The 'it' must be anaphoric and so refer to the *feature*, likely a property or trope.

[61] Caveat: if Lewis (1979) is correct, we can believe a limited range of properties (e.g., being in a world where p is true). This view is controversial, but even if it were right, my point would survive: we cannot believe the analogues of the properties we can admire.

(38) #I believe being supported by strong evidence in a proposition.

This follows from the Hairetic Theory of Belief Formation, because properties aren't possibly good answers to questions. Thus, there is no option set that is something like {believing F-ness: F-ness is a property propositions can have}. That's just not the kind of attitude belief is.

We have nearly all the tools we need to explain the data in section 2 about the non-doxastic attitudes. Suppose that S is trying to figure out what properties to admire; {admiring F_1-ness in a G_1, admiring F_2-ness in a G_2, ...} are their options. In this context, if they decide to admire F-ness, then—I claim—they admire whatever *has* F-ness. So, e.g., I can admire Malala Yousafzai's bravery and dedication. Indeed, I can admire bravery and dedication in activists more generally (or in "an activist"). It seems to follow from that that I admire brave and dedicated activists *for* their bravery and dedication. After all, it sounds weird to say 'I admire bravery and dedication in an activist, but I don't admire those activists who are brave and dedicated for their bravery and dedication'. But if I admire someone for something, it follows that I admire them. After all, it sounds weird to say 'I admire her for her bravery and dedication, but I don't admire her'. Of course, these steps could be questioned, but they are at least intuitive. In short, admiring properties means we admire the people that have those properties *for*—minimally, insofar as they have—that property. This does not require that we admire them "overall". I can admire my nemesis's dedication, and admire them for their dedication, even though I think they're bad overall so that I wouldn't admire them overall, i.e., all-things-considered. But not all admiration is all-things-considered.

Thus, admiring giving large amounts of time to help people who need it entails admiring whoever gives large amounts of time to help people who need it. Of course, if two of the options S had considered were admiring giving large amounts of time to help people who need it in order to impress a crush and admiring giving large amounts of time to help people who need it simply in order to help those people, S would likely admire the latter and not the former, and so the entailment wouldn't go through. But when the choice is the simpler one, the entailment does go through. That's a reflection of the option sensitivity I began this section discussing.

Here is the analogue of Hairetic Theory of Belief Formation I have argued for for admiration:

> Hairetic Theory of Admiration Formation. S comes to admire o given her option set $\Phi = \{\phi_1, ..., \phi_n\}$ only if either some $\phi_i \in \Phi$ is admiring o and S chooses ϕ_i, or some $\phi_i \in \Phi$ is admiring F-ness in a G, o is an F G, and S chooses ϕ_i.

I have focused on Hairetic Theory of Admiration Formation in particular, though I

think the arguments I have developed for it will apply *mutatis mutandis* to some of the other non-doxastic attitudes of interest like hope, hatred, and anger. Take hope: I hope for excellence in the paper I'm starting, and I fear unoriginality in it. When I hope for excellence, I hope *that* it's excellent, and when I fear unoriginality in it, I fear *that* it'll be unoriginal. I hope for my children's happiness. So I can hope for properties. Of course, I can't hope properties. But as I argued earlier in section 2, we can hope for them, even hope for them *in* things. So the same arguments will apply to hope.

I am at last ready to explain the phenomena from section 2. First, why is some non-doxastic modal tracking possible? It is possible because option sets for decisions about admiration include admiring *properties*. Because different things have different properties in different worlds, agents will admire different things in different worlds in virtue of admiring the same properties in those worlds. Thus, I can hate whoever was willing to damage my car without apologizing; and so I can also be such that, if A did it, I hate them, and if B did it, I hate them instead. This contrasts with belief, because, recall, we cannot believe properties, just propositions. Thus we cannot believe propositions *in virtue of* believing properties. This is how we can explain the contrast between doxastic and non-doxastic attitudes.

That was not the only strange aspect of the phenomena in section 2, though. Remember that, though non-doxastic attitudes can be modally unstable, it seemed that they cannot be mediated by just any properties, in particular by properties like $\lambda x.x$ is someone I ought to admire.

Here's my answer. I want to admire people that I ought to admire. But I don't admire people *for* being such that I ought to admire them. I don't admire that feature in them. I would only admire the features in them that would *make* them such that I ought to admire them. So I'd admire altruism in a person, for example. Being a person whom one ought to admire is not itself a property that it is fitting to admire in a person, and since this is pretty clear to me, I can't bring myself to admire it. But unless I have the correct view of all and only those features that make a person someone I ought to admire, I won't admire whomever I ought to admire. Nor can an individual come to admire whatever *in fact* makes people most admirable, by thinking of it that way; if the alternatives in a choice about whom to admire are things like helping the needy, creating large companies, etc., with roughly those modes of presentation, then attempting to choose properties to admire this way would be a violation of Clarity.[62] That's why (19) and its analogues sound bad.

Similarly, and finally, why does it sound questionable to say the following?

[62] It's an important question what determines which properties help constitute these option sets when they do—with belief, the answer was 'they are possibly good answers to questions the agent's trying to answer'. I suspect properties have to be presented in a way that would make it intelligible how they would make things admirable. But that's speculation; I only note that we'll need *some* explanation, since I can't admire properties like the one denoted by 'the property that makes the people my friends admire admirable'. Everyone needs to explain that, so I leave it to future work.

(39) I admire whoever my friend Anna admires, though I have no idea who she does admire or what she finds admirable.[63]

It sounds bad because it is difficult to admire being admired by Anna, because that property itself is not admirable; it just correlates with admirableness, if she's reliable. When we admire whatever is F in virtue of admiring F-ness, we need to be able to admire F-ness in the first place. But it is very difficult to admire properties in people when we don't think those properties themselves are admirable.

The explanations here of the phenomena in section 2 are general across attitude-types that work similarly to admiration. I conjecture, e.g., hate and hope are both like that. The arguments are simple to construct. That this conception of non-doxastic attitudes is natural and allows us to explain the data from sections 1 and 2 more broadly constitutes one more argument in favor of the conception itself.

This completes the presentation and defense of my hairetic theory of attitude-formation. The idea was to apply the solution worked out for choice in section 3 to doxastic and non-doxastic attitudes, and derive the difference between them from the differences in the option sets corresponding to those attitudes. That it can explain the data in section 2 is yet more reason to believe the hairetic theory is right.

7 Conclusion

The puzzle with which I began was this:

- Why can we not rationally believe H by fixing its reference in the way we did? Why can't we have non-constant modally tracking belief states in cases where externalist content-determining mechanisms don't play a role?
- Why *can* we have non-constant modally tracking non-doxastic states, like for hope and admiration?
- But why can't they track properties like $\lambda x.x$ is admirable and $\lambda x.x$ is admired by my friend?

To account for the phenomena, I showed that a similar thing happens with choice. The principle Clarity seems to explain it in that situation. I then argued that attitude-formation more generally ought to be conceived of as choices, that is (in my terminology) *hairetically*. Partly there are independent arguments for this conclusion, namely that data involving these attitudes can be explained if some hairetic theory of their formation is correct. But

[63] There is a reading of the first part of this sentence that sounds fine, where the speaker is confident they have the same values as Anna does. I don't mean to discuss that reading.

it's also important that hairetic theories of the formation of these attitudes can, with Clarity, explain the phenomena from sections 1 and 2. Specifically, the hairetic theory of belief formation entails that we cannot believe properties but rather only things that might be good answers to some questions we ask ourselves. That explains why we cannot have non-constant modally tracking doxastic states, which combines with Clarity to explain why we cannot rationally believe *H* in the relevant circumstances. Finally, the *differences* between these attitudes from sections 1 and 2 simply arise from differences in their respective option sets. In light of the ability of the hairetic theory of doxastic and non-doxastic attitudes to explain so much, I think we should believe that it is true.

Appendix: Doxastic Modal Tracking and Conditionals

In section 1, I exhibited linguistic evidence that doxastic modal tracking is only possible when an externalist content-determining mechanism like names whose reference was fixed by description make it possible. Thus, (3) and (4) sound bad, while (5) sounds fine. The problem is that there are conditionals that are at least apparently similar to (3) and (4) that sound fine, but wouldn't be interpreted with any relevant externalist content-determining mechanism. Since the conditionals are pretty important to establish Stability Constraint for Belief as a true generalization, in this appendix, I'm going to go through these other problematic conditionals with 'believe' and show why they don't present any problems.

I used this bad-sounding conditional to argue for the Stability Constraint for Belief:

(3) #If there was a unique individual who wrote the *Iliad* and the *Odyssey*, then I believe there was, and if not, then I believe there wasn't.

Not all such conditionals sound bad, even when the speaker would seem to violate the Stability Constraint for Belief. Suppose I say of some friends:

(16) If they stole my lunch, then I think they're in big trouble!

This sounds fine. So it might seem that the Stability Constraint for Belief isn't true. Here I'll present an alternative interpretation of conditionals like (16) and argue that that interpretation better accounts for the data. (This will also explain why (4) sounds a bit better than (3) does.)

The surface form of the conditionals is ⌜if σ, I Ψ o⌝, where, so far, o is a proposition. To this surface form corresponds different underlying logical forms, depending on how we resolve scope ambiguities. According to a proposal of Schlenker (2005)'s (originally considered and rejected by Lewis (1973)), an 'if'-clause functions as a definite description.

In other words, ⌜if σ⌝ means roughly what ⌜the (closest) σ-situations⌝ means.[64] Now suppose we have a sentence of the form ⌜S Ψ-s that the F is G⌝, where ⌜Ψ⌝ is an attitude verb. That is famously ambiguous between *de dicto* and *de re* readings:

(40) a. S Ψ-s that the F is G.

b. Sam believes that the star of *Top Gun* is tall.

(41) a. S Ψ-s of the F that it is G

b. Sam believes of the star of *Top Gun* that he not is tall.

(40b) can be true because Sam thinks any male movie star is tall, even though (41b) is true because the guy he sees in real life (who is Tom Cruise) is short.

Let's assume Schlenker's right about 'if'-clauses being definite descriptions as I described. Then they should exhibit the same ambiguity with respect to attitude verbs. And because we can *also* interpret the attitude verb as being part of the consequent, we have an at least three-way ambiguity. So we have the following (scope descriptions are determined by the conditional-as-definite description):

(16) a. *Narrow scope* = The closest situation s in which they stole my lunch is such that I, in s, believe they're in big trouble.

b. *Wide scope*, de dicto = I believe that the closest situation s in which they stole my lunch is such that they're in big trouble in s.

c. *Wide scope*, de re = I believe of the closest situation s in which they stole my lunch that they are in big trouble in s.

So in principle we have these three readings available for (16). (16a) would be bad for me, while (16b) and (16c) would be fine. So, do we have reason to interpret it as (b) or (c) rather than (a)?

We do. Notice (16) and the following are assertible in the same circumstances:

(42) If they stole my lunch, then they're in big trouble.

That is hard to understand if (16) has the form (16a): the consequents seem to be about different things, namely the speaker's mental state in (16) and whether the friends are in trouble in (42). Yet it makes sense on either (b) or (c). That's because ⌜σ⌝ and ⌜I think that σ⌝ are usually assertible in the same circumstances.

Notice also that these sound bad:

[64] For similar ideas with different implementations, see Haiman (1978) and Bittner (2001).

(43) a. ???If they stole my lunch, then I think they stole it.

b. ???If they stole my lunch, then I think they took it.

With utterances of the form ⌜if σ, I believe that σ'⌝, it seems that σ cannot entail σ'. (3) was another example of this. That's puzzling if we can interpret these as (16a) does. If narrow scope interpretations are available for sentences where σ doesn't entail σ', we have no explanation of why it sounds bad when it does. But if we have to interpret such sentences with 'believe' as either wide scope, *de dicto* or wide scope, *de re*, this makes sense: they would then be *trivial*; *everyone* believes the closest σ-situation is a σ-situation (or a σ'-situation, where σ clearly entails σ').

So I think we have reason to interpret (16)-like conditionals as wide scope, either *de dicto* or *de re*. I'll come back to which soon. What are we *doing* when we assert conditionals like that? I say that we are *commenting on* situations. Consider:[65]

(44) Suppose a wolf came in but we missed it somehow. In that case I think we're in a lot of trouble!

(44) is acceptable. But 'I think' does not refer to an attitude we have *within* the situation, but is a commentary *on* the situation. Consider, e.g., a counterfactual:

(45) If evolution had gone a little differently and the eliminative materialists were right, I think we would be a little less interesting than we in fact are. That would have been a very sad situation!

Two things are notable here. First, if 'think' were "part" of the consequent, it should be 'would think', which wouldn't make sense, given the antecedent. Second, 'I think' seems to be doing the *same thing* in (45) it does in (16). If that impression is correct, since 'I think' in (45) cannot be understood as being about what I think *in* the counterfactual situation, we should think (16) isn't about that, either. (16) seems to me to express what the speaker thinks *about* the counterfactual situation.

In principle either (16b) or (16c) could allow a speaker to do this. But wide scope, *de dicto* interpretations have problems when it comes to other attitude verbs that entail belief or lack thereof. Consider:

(46) If my friend left without saying "goodbye", I'm surprised that she did.[66]

The wide scope, *de dicto* interpretation of (46) is:

[65] See the modal subordination literature, e.g., Roberts (1989) and Stone (1999).

[66] This example more or less comes from Blumberg and Holguín (forthcoming), though as an argument against wide scope interpretations it comes from Drucker (2019).

(46b) I'm surprised that if my friend left without saying "goodbye", that she left without saying "goodbye".

But that's *not* what (46) is really saying, because the speaker is not saying that she is surprised by a tautology. To handle examples like that, we need to appeal to wide scope, *de re* readings. That gives the following interpretation of (46):

(46c) I'm surprised *of* the closest situation *s* such that my friend left without saying "goodbye", that she left without saying "goodbye" in *s*.

This interpretation does not have the speaker express surprise at a tautology, but rather surprise at a situation that might be non-actual. But it is perfectly appropriate to find such a situation surprising! The speaker's friend presumably acts quite uncharacteristically in it. This is not very different from being surprised that some character in a piece of fiction acts uncharacteristically. No one need be surprised *in* the fiction, or surprised about a tautology. Thus, wide scope, *de re* readings exist to play the role of the commentary conditionals we've been looking for.[67]

If Schlenker's even roughly right, then the different readings will always be syntactically available. But they will not all be natural, or what a speaker actually intends to communicate. As always, to disambiguate, we need to rely on a combination of world knowledge, incipient grasp of the speaker's intentions, and what would be a reasonable or rational thing to want to communicate.

Finally, I would like there to be a difference between conditionals like (3) and (16), where the narrow scope interpretations are unavailable (because, I think, obviously false), and conditionals like (15), where they are available:

[67] A referee objects that the following sort of example is fine:

(47) If my friends stole my lunch, then I am worried that they don't have enough money to buy lunch for themselves. But I don't think they did steal my lunch, so I'm not worried about anything.

But, they continue, shouldn't the last sentence of (47) sound bad? On my account, the speaker would still be worried *of* the (closest) situation in which my friends stole my lunch that they don't have enough money. In response, note first, usually when we say we are worried or not, we restrict the worries we have in mind to those that we take to be relatively pressing. Thus, I am worried of the closest situation in which my dog has been faking his affection that he doesn't love me, but I would still say I'm not worried about anything about my dog, because that is an incredibly unlikely situation to be actual. Second, 'think' is a neg-raising verb; ⌜I don't think that p⌝ implies, semantically or pragmatically, what ⌜I think that $\neg p$⌝ does. Typically, ⌜I am worried that p⌝ suggests that p and $\neg p$ are open for the speaker. (Consider: 'I am worried that it'll rain'. It sounds odd if the speaker knows it will or it won't.) But if the speaker says they don't think their friends stole anything, it's not open for them that they did, and thus they can't worry that they did. Compare:

(48) ...But I don't know whether they did, so I'm not actually worried about anything.

This sounds much worse to me than (47) does. (47)'s acceptability, then, is no problem for the account I give in this section of (16).

(15) If having children would bring me the most happiness, I hope that I have children.

I essentially have already argued for this. Recall that in that context, (14)'s logical form seems to be or entail '[every x: x is either that John have children or that John not have children](x would bring John the most happiness $\supset$ John hopes for x)'. And this logical form, along with (18) ('if having children were the thing that would bring John the most happiness, John would still hope for whatever would bring him the most happiness'), entails the narrow scope reading of (15). Even better, we can't use the same trick with belief, since (1) ('I believe whichever of Homerism and anti-Homerism is true') sounds bad—indeed, is false, I think. That is why we should, and I think do, interpret them differently.

References

Aloni, Maria and Jacinto, Bruno. 2014. "Knowing Who: How Perspectives and Context Interact." In Franck Lihoreau and Manuel Rebuschi (eds.), *Epistemology, Context, and Formalism*, 81–107. Dordrecht: Springer.

Anand, Pranav and Hacquard, Valentine. 2013. "Epistemics and Attitudes." *Semantics & Pragmatics* 6:1–59.

Anscombe, G. E. M. 2000. *Intention*. Cambridge, MA: Harvard. Originally published 1957.

Arpaly, Nomy and Schroeder, Timothy. 2014. *In Praise of Desire*. Oxford, UK: Oxford.

Beaver, David I. and Clark, Brady Z. 2008. *Sense and Sensitivity: How Focus Determines Meaning*. Malden, MA: Wiley-Blackwell.

Bittner, Maria. 2001. "Topical Referents for Individuals and Possibilities." In Rachel Hastings, Brendan Jackson, and Zsofia Zvolensky (eds.), *Proceedings of SALT*, volume XI, 36–55. Ithaca, NY: Cornell.

Blumberg, Kyle and Holguín, Ben. forthcoming. "Embedded Attitudes." *Journal of Semantics* .

Boër, Steven E. and Lycan, William G. 1975. "Knowing Who." *Philosophical Studies* 28:299–344.

Boyle, Matthew. 2011. "'Making up Your Mind' and the Activity of Reason." *Philosophers' Imprint* 11:1–24.

Bratman, Michael E. 2009. "Intention, Belief, Practical, Theoretical." In Simon Robertson (ed.), *Spheres of Reason: New Essays in the Philosophy of Normativity*. Oxford, UK: Oxford.

Braun, David. 1998. "Understanding Belief Reports." *Philosophical Review* 107:555–595.

Chierchia, Gennaro and Turner, Raymond. 1988. "Semantics and Property Theory." *Linguistics and Philosophy* 11:261–302.

Davidson, Donald. 1963. "Actions, Reasons, and Causes." *Journal of Philosophy* 60:685–700.

De Sousa, Ronald. 1971. "How to Give a Piece of Your Mind: Or, the Logic of Belief and Assent." *Review of Metaphysics* 25:52–79.

Dixon, Robert M. W. 2005. *A Semantic Approach to English Grammar*. New York: Oxford. Second edition.

Donnellan, Keith. 1977. "The Contingent *A Priori* and Rigid Designators." In Peter French, Theodore E. Uehling, Jr, and Howard K. Wettstein (eds.), *Midwest Studies in Philosophy II: Studies in the Philosophy of Language*, 12–27. Minneapolis, MN: University of Minnesota Press.

Dretske, Fred. 1981. "The Pragmatic Dimension of Knowledge." *Philosophical Studies* 40:363–378.

Drucker, Daniel. 2019. "Policy Externalism." *Philosophy and Phenomenological Research* 98:261–285.

Dummett, Michael. 1973. *Frege: Philosophy of Language*. London: Duckworth.

Edgington, Dorothy. 1995. "On Conditionals." *Mind* 104:235–329.

Evans, Gareth. 1982. *The Varieties of Reference*. Oxford: Oxford University Press. Edited by John McDowell.

Forbes, Graeme. 2000. "Objectual Attitudes." *Linguistics and Philosophy* 23:141–183.

Friedman, Jane. 2013. "Question-Directed Attitudes." *Philosophical Perspectives* 27:145–174.

Greaves, Hilary. 2013. "Epistemic Decision Theory." *Mind* 122:915–952.

Grice, H. P. 1971. "Intention and Uncertainty." In *Proceedings of the British Academy*, volume 57, 3–19. London, UK: Oxford.

Haiman, John. 1978. "Conditionals Are Topics." *Language* 54:564–589.

Hamblin, C. L. 1958. "Questions." *Australasian Journal of Philosophy* 36:159–168.

Harman, Gilbert. 1976. "Practical Reasoning." *The Review of Metaphysics* 29:431–463.

Hedden, Brian. 2012. "Options and the Subjective *Ought*." *Philosophical Studies* 158:343–360.

Heim, Irene. 1992. "Presupposition Projection and the Semantics of Attitude Verbs." *Journal of Semantics* 9:183–221.

Heim, Irene and Kratzer, Angelika. 1998. *Semantics in Generative Grammar*. Malden, MA: Blackwell.

Hieronymi, Pamela. 2006. "Controlling Attitudes." *Pacific Philosophical Quarterly* 87:45–74.

Holliday, Wesley H. 2015. "Fallibilism and Multiple Paths to Knowledge." In Tamar Szabò Gendler and John Hawthorne (eds.), *Oxford Studies in Epistemology*, volume 5, 97–144. Oxford: Oxford.

Jacobson, Pauline. 1995. "On the Quantificational Force of Free Relatives." In Emmon Bach, Eloise Jelinek, Angelika Kratzer, and Barbara Partee (eds.), *Quantification in Natural Languages*, 451–486. Dordrecht, Netherlands: Kluwer.

Jerzak, Ethan. 2019. "Non-Classical Knowledge." *Philosophy and Phenomenological Research* 98:190–220.

Joyce, James M. 1998. "A Nonpragmatic Vindication of Probabilism." *Philosophy of Science* 65:575–603.

Kaplan, David. 1968. "Quantifying In." *Synthese* 19:178–214.

King, Jeffrey C. 2002. "Designating Propositions." *Philosophical Review* 111:341–371.

Korsgaard, Christine M. 2009. "The Activity of Reason." *Proceedings and Addresses of the American Philosophical Association* 83:27–47.

Kripke, Saul. 1972. "Naming and Necessity." In Donald Davidson and Gilbert Harman (eds.), *Semantics of Natural Language*, 253–355, 763–9. Dordrecht: D. Reidel. Revised edition published in 1980 as *Naming and Necessity* (Harvard University Press, Cambridge, MA).

—. 2011. "Frege's Theory of Sense and Reference: Some Exegetical Notes." In *Philosophical Troubles*, 254–291. Oxford, UK: Oxford.

Lackey, Jennifer. 2007. "Norms of Assertion." *Noûs* 41:594–626.

Lewis, David. 1973. *Counterfactuals*. Oxford: Blackwell.

—. 1979. "Attitudes *De Dicto* and *De Se*." *Philosophical Review* 88:513–43.

—. 1996. "Elusive Knowledge." *Australasian Journal of Philosophy* 74:549–67.

Mates, Benson. 1952. "Synonymity." In Leonard Linsky (ed.), *Semantics and the Philosophy of Language*. Urbana, IL: Illinois.

Merricks, Trenton. 2009. "Propositional Attitudes?" *Proceedings of the Aristotelian Society* 109:207–332.

Moltmann, Friederike. 2003. "Propositional Attitudes Without Propositions." *Synthese* 135:77–118.

Montague, Michelle. 2007. "Against Propositionalism." *Noûs* 41:503–518.

Nebel, Jake. 2019. "Hopes, Fears, and Other Grammatical Scarecrows." *Philosophical Review* 128:63–105.

Peacocke, Christopher. 1998. "Conscious Attitudes, Attention, and Self-Knowledge." In Crispin Wright, Barry Smith, and Cynthia MacDonald (eds.), *Knowing Our Own Minds*. Oxford, UK: Oxford.

Perry, John. 1980. "A Problem About Continued Belief." *Pacific Philosophical Quarterly* 61:317–332.

Prior, A. N. 1971. *Objects of Thought*. Oxford: Oxford University Press.

Quine, W. V. 1981. "Intensions Revisited." In *Theories and Things*, 113–123. Cambridge, MA: Harvard.

Quine, W. V. O. 1948. "On What There Is." *Review of Metaphysics* 2:21–38.

Roberts, Craige. 1989. "Modal Subordination and Pronominal Anaphora in Discourse." *Linguistics and Philosophy* 12:683–721.

—. 2012. "Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics." *Semantics and Pragmatics* 5:1–69. Written in 1996.

Rooth, Mats. 1992. "A Theory of Focus Interpretation." *Natural Language Semantics* 1:75–116.

Salmon, Nathan. 1986. *Frege's Puzzle*. Cambridge, MA: MIT.

Salmon, Nathan and Soames, Scott. 1988. *Propositions and Attitudes*. Oxford: Oxford.

Schaffer, Jonathan. 2007. "Knowing the Answer." *Philosophy and Phenomenological Research* 75:383–403.

Schaffer, Jonathan and Szabó, Zoltán Gendler. 2014. "Epistemic Comparativism: A Contextualist Semantics for Knowledge Ascriptions." *Philosophical Studies* 168:491–543.

Schlenker, Philippe. 2005. "Conditionals as Definite Descriptions." *Research on Language and Computation* 2:417–462.

Shah, Nishi and Velleman, J. David. 2005. "Doxastic Deliberation." *Philosophical Review* 114:497–534.

Soames, Scott. 1989. "Semantics and Semantic Competence." *Philosophical Perspectives* 3:575–596.

Stalnaker, Robert C. 1984. *Inquiry*. Cambridge: MIT.

Stone, Matthew. 1999. "Reference to Possible Worlds."

Swanson, Eric. 2012. "Propositional Attitudes." In Claudia Maienborn, Klaus von Heusinger, and Paul Portner (eds.), *Semantics: An International Handbook of Natural Language Meaning*, volume 2, 1538–1561. Berlin, Germany: Mouton de Gruyter.

Theiler, Nadine, Roelofsen, Floris, and Aloni, Maria. forthcoming. "A Uniform Semantics for Declarative and Interrogative Complements." *Journal of Semantics* .

Šimík, Radek. forthcoming. "Free Relatives." In Daniel Gutzmann, Lisa Matthewson, Cécile Meier, Hotze Rullmann, and Thomas Ede Zimmermann (eds.), *The Companion to Semantics*. Wiley.

Williams, Bernard. 1973. "Deciding to Believe." In *Problems of the Self*, 136–151. Cambridge: Cambridge.

Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford: Oxford.

—. 2013. *Identity and Discrimination*. Oxford: Wiley-Blackwell. 2nd Edition.

Wright, Sarah. 2014. "The Stoic Epistemic Virtues of Groups." In Jennifer Lackey (ed.), *Essays in Collective Epistemology*, 122–141. Oxford, UK: Oxford.

Yalcin, Seth. 2018. "Belief as Question-Sensitive." *Philosophy and Phenomenological Research* 97:23–47.