

# 12

## MODELING WITH HYPERPLANS

*Seth Yalcin*

A key gizmo in Gibbard's formal model of normative judgment is the thing he used to call a "complete system of norms" (Gibbard 1986, 1990), and which, by around Gibbard (2003), evolved into the *hyperplan*. In this paper, I want to ask what hyperplans are, and ask how best to use them in modeling normative thinking.

One part of this paper is exegetical. There is perhaps less than universal agreement in the literature on how Gibbard's systematizing with hyperplans works exactly. I offer a take.

The other part is exploratory. I think there is a way of theorizing with hyperplans that is not quite Gibbard's way, but which is also expressivistic, and which is worth looking at. I have tried to say so in Yalcin (2012, 2018), but here I offer a more focused development. I will call (my take on) Gibbard's package of views about how to model with hyperplans Plan A. I spend the first section of the paper setting Plan A out. The alternative I will set out, Plan B, is the topic of the sections after that. Even if you don't leave the paper preferring Plan B to Plan A, I hope you'll find the contrast clarifying. In separating these two ways of theorizing with hyperplans, really I'm trying to bring out two rather different ways of conceptualizing expressivism.

### 1. Plan A: Gibbard's Model

#### *1.1. Easing into the Picture*

Hyperplans ultimately figure in an abstract model of what it is for the idealized rational thinker-planner to be normatively opinionated. To get into the model, a helpful starting point is the sort of picture of agents we find in a decision-theoretic perspective, where we suppose that agents have (1) a take on what the world is (probably) like, representable

formally by a probability measure over a space of possible outcomes, and (2) some preferences, representable as a partial ordering of possible outcomes. Or if we are satisfied to ignore distinctions between degrees of confidence, as I will be for now, we can replace the probability measure with a set of possible situations, the situations that might obtain for all the agent believes—the agent’s *doxastic alternatives*. A textbook modeler in this vein will use it to say that a rational agent is the sort of being that generally acts in ways that would realize her preferences, at least relative to possible situations compatible with how she takes things to be. That gives an initial skeletal picture about how belief and desire connect to action, one not totally alien from the way ordinary belief and desire talk is invoked to explain behavior.

Saying only this much still leaves a lot of wiggle room, of course, when it comes to the question how exactly to interpret natural language talk of mental states—ascriptions of belief, desire, and the like—using elements of the model. We informally use this language when explaining in the first place what the model is supposed to be representing, but we could be more systematic. So, we could add to the story, stating idealized truth-conditions for ordinary ascriptions of belief and desire using the concepts of this model. The model has a component approximating what the folk call “belief” and a component approximating what the folk call “preference.” We could invoke these components in stating truth-conditions for belief and desire talk. Take for instance:

- (1) Holmes thinks it is not too late to catch the train.
- (2) Holmes wants to get on the train.

The usual move is to say that when (1) is true, that’s to do with the agent’s doxastic alternatives: it is true just in case none of Holmes’s doxastic alternatives are ones where it is too late to catch the train.<sup>1</sup>

For (2), it is initially tempting to say that its truth is to do just with the agent’s preference ordering, as the truth of (1) is just to do with the agent’s doxastically open possibilities. But on reflection, wanting is a more subtle thing than that: what one wants depends partly on what one thinks the world is like, and not just on which possible states one prefers to others. I want to sue the man whose dog bit me, but this want of mine cannot necessarily be read off my preference ordering alone. It’s not that I prefer situations where I sue the man to all others—after all, in the possible situations I most prefer, I am never bitten by a dog at all. It seems like wanting a possible outcome is more like “preferring it to certain relevant alternatives, the relevant alternatives being those possibilities that the agent believes will be realized if he does not get what he wants” (Stalnaker 1984, 89). If that is right, then whether an agent wants something is a function of both their preference ordering and their doxastic state. Wanting cannot be modeled with a preference ordering alone, but neither does it map

<sup>1</sup> So it goes in the familiar tradition of possible worlds modeling descending from Hintikka (1962) and running prominently through theorists like Stalnaker (1976, 1984) and Lewis (1979).

to its own special feature of the model, as belief basically does. When you say somebody wants something, you're saying that their preference ordering and their doxastic state, taken as a pair, satisfy a certain complex property—one in the direction of what Stalnaker suggests, I'd claim, though the details don't matter right now.<sup>2</sup>

In the last paragraph, I mean to draw out the point that although states of preference are basic to our model, some of the key pieces of ordinary language we use to describe an agent's preferences—desire ascriptions, ascription of want—might well have a more subtle connection to the model than one might offhand suppose. We have to distinguish concepts from the model (like preference orderings) from ordinary language notions (like wanting), and we have to allow for the possibility that they may be connected by something other than a straight line. This will be good to have in mind as we turn to normative judgment.

### 1.2. Adding Normative Judgment

Now suppose we want to extend our model, adding a formalization of what it is to have normative views. I want to approach this via a concrete example, like:

- (3) Holmes thinks he ought to pack.

Given the sort of model we are assuming, we can ask: what does a model of Holmes's state have to be like to render this true? What features of our model, if any, correspond to what this sentence says?

An unfussy normative realism would favor the view that there is nothing fundamentally new to say here. Holmes's thinking he ought to pack is not deeply different than his thinking it's not too late to catch the train. Both are about the way he takes the world to be, about what he thinks the facts are.<sup>3</sup> On this view, to see whether our model of Holmes reveals that he thinks he ought to pack, we just check whether his doxastic alternatives are all possibilities where he ought to pack. Possible states of the world fix facts such as Holmes's being required to pack, on this view—just as much as they fix facts about train schedules.

Alternatively, we can say that's wrong. Holmes's thinking he ought to pack is not just another one of his views about how the world is. We cannot discover whether he is in this state just by looking at the doxastic alternatives he keeps open. There is something like an official list of classic motivations for rejecting realist-style truth-conditions for (3), much of it descending from Hume (usually it includes: the alleged metaphysical queerness of the realist's normative properties, the direction of fit of normative judgment, its alleged internal link to motivation, the open question argument, and the supposed intractability of normative disagreement), but our interest is just in getting some interesting alternative analyses of (3)

<sup>2</sup> Heim (1992), Levinson (2003), and Lassiter (2011) contain more on the semantics of 'wants'.

<sup>3</sup> Perhaps *modulo* whatever *de se* or self-locating elements we might think infect these thoughts.

on the table, Gibbard's especially. Gibbard wants to say that the truth of (3) turns on features of Holmes's mental economy beyond his doxastic alternatives. What features?

Well, so far our model of an agent includes only a set of doxastic alternatives and a preference ordering, so either ought-thoughts are tied up somehow with preference, or else we need to expand our model of Holmes's mental economy. Consider briefly the first option. An idea in this vein would be to basically identify ought-thoughts with wants: we could say that states of *thinking it ought to be that p* are tantamount to states of *wanting that p*. Both kinds of ascription serve to place constraints on the agent's preference ordering (and the agent's doxastic state, too, if such states really are equivalent to states of wanting). Thus *ought*-thoughts are not purely doxastic—we could say they are “preference-laden.” Carnap and Russell seemed to favor this kind of analysis—they suggested that normative statements express wishes (see Carnap 1935, 24; Russell 1935, 236–7). This approach recommends the idea that (3) serves to tell us about what Holmes wishes or wants.

The evident trouble is that it seems like (3) can be true even when it's also true that Holmes prefers not to pack. What I think ought to be the case and what I want, it seems obvious, can come apart. Maybe desiring something is a way of valuing it, but there clearly are ways of valuing, or evaluating, things that are not just ways of preferring them; so anyway I'll assume without argument. Should we nevertheless try to say that if (3) is true, then really Holmes *does* want to pack in some maybe very attenuated or nuanced sense of *want*? Or, less committally, that the relevant state of mind is “desire-like”? Such a view could be developed,<sup>4</sup> but we mean to get to Gibbard, and his view seems to me not helpfully put this way, so I will just take it our model must go beyond doxastic states and preference orderings, beyond belief and desire as we have them now modeled, to bring normative attitudes into its ken.

Now eyeing directions for expansion, one possibility would be to bring in feelings. There is the notorious emotivist strand of expressivism, prominent in Ayer (1936), which ties normative states of mind to emotional states. This view is in the ballpark of saying that (3) reports that Holmes has good vibes about packing. I set this idea aside too in order to focus on a different idea, the idea of bringing in states of mind to do with planning and intention—the idea Gibbard favors and develops in Gibbard (2003) particularly.

### 1.3. Planning

A good first step into it is to follow Bratman (1987). Against philosophers like Anscombe and Davidson, we don't try to explain states of intention away adverbially, analyzing them in terms of the notion of acting intentionally, the latter then analyzed as a matter of the right fit obtaining between the action and the agent's beliefs and desires. Instead, we think

<sup>4</sup> Hare offered a sophisticated version of this kind of approach, holding that moral claims express preferences the agent takes in a certain sense to be universalizable (Hare 1981). One often sees expressivism characterized as a view that holds that normative statements express desire-like states of mind (e.g., Ridge 2014; Schroeder 2009). Often what is intended here is a similarity in “direction of fit.”

of intentions as pieces of plans agents have, and we take it planning states of mind are their own thing—they do not admit of reduction to belief, preference, or some admixture. They have respectable standing of their own, on par with those other states. Our formal model of agents needs expanding to make room for them, take it. We could, following Bratman, see the addition of intention to the picture as rendering our model less idealized, bringing in a state of mind whose *raison d'être* is partly the fact that we don't have unlimited mental resources for reevaluating the pros and cons of the options we face as the facts change.

So we want to expand the model, adding planning. Here is where hyperplans enter the picture. Gibbard's approach is to elaborate our model of doxastic states, interleaving in a planning dimension. Previously, we had it that a doxastic state is a set of possibilities—a set of ways that the world could factually be, for all the agent believes. Now following Gibbard we say that a doxastic state is given by a set of pairs of a possibility and a hyperplan. Let me explain what formally a hyperplan is, and then try to say how by adding these things to our model Gibbard means to clarify planning thinking.

On the formal side:

A *hyperplan*, we can stipulate, covers any occasion for choice one might conceivably be in, and for each alternative open on such an occasion, to adopt the plan involves either rejecting that alternative or rejecting it. In other words, the plan either forbids an alternative or permits it.

(Gibbard 2003, 56)

For any possible occasion of choice, a hyperplan settles some nonempty subset of the options available on that occasion—intuitively, the ones deemed permissible by the hyperplan. We can take hyperplans to be functions on centered worlds.<sup>5</sup> Fed a centered world  $c$ —an “occasion of choice”—where the agent centered at  $c$  faces a set of options  $O_c$  (the options fixed by the centered world  $c$ ), the hyperplan  $h$  outputs some nonempty subset of  $O_c$ —intuitively, those options deemed permissible by  $h$ . What formally are options? We could think of an option as a set of centered worlds, the set of centered worlds where the option is realized.

A hyperplan is not the sort of thing tied to any one particular agent—it is not a complete contingency plan for some one particular planner. Rather, it settles what is okay to do for *every* planner, actual or possible, and for any possible situation any planner could find herself in. Hyperplans are defined on arbitrary centered worlds.

Let's say a *fact-plan alternative* is a pair of a centered world and a hyperplan. A Gibbardian plan-laden belief state is a set of fact-plan alternatives. Thus, Gibbard is thinking of plan-laden belief as cutting a strictly richer space of alternatives than we previously had it. It is

5 “A *situation*  $s$  is a triple  $\langle w, i, t \rangle$  of a world  $w$ , an agent  $i$  in  $w$ , and a time  $t$  at which agent  $i$  in world  $w$  has a choice of what to do. For each such situation  $s$ , there is a set  $a(s)$  of *alternatives*. These are maximally specific acts open to person  $i$  at time  $t$  in world  $w$ . A *hyperplan*  $p$  assigns to each situation  $s$  a non-empty subset  $p(s)$  of the alternative set  $a(s)$ ” (Gibbard, 2003, 100).

helpful to see him as building on Lewis (1979). First, he is taking on board Lewis's main idea, that to model belief *de se* in full generality we do better to think of doxastic alternatives as centered worlds rather than uncentered worlds. Then going beyond Lewis and considering the yet-richer space of fact-plan alternatives, he proposes that subsets of this space can well represent our states of mind that combine belief and decision.

Since a fact-plan alternative fixes a centered world, a set of fact-plan alternatives can from a modeling perspective do everything a set of centered worlds can do. Therefore, Gibbard's model of belief is at least as robust as Lewis's. Indeed, Gibbard will model ordinary (if ideal) "prosaically factual" belief essentially along Lewis's lines, as a matter of what centered worlds figure in the fact-plan alternatives the agent leaves open. For example, let  $P$  be Holmes's plan-laden belief state, a set of fact-plan alternatives. Gibbard will say that (1)—an ascription to Holmes of the prosaically factual (if self-locating) belief that isn't not too late to catch the train—is true just in case:

For all  $\langle c, h \rangle \in P$ , it isn't too late to catch the train at  $c$

You can see that the hyperplan component is an idle wheel. That wheel hits the pavement only when we come to planning and normative judgment. Gibbard says that you have views about how things are and views about what to do. Your views about what to do are not settled by your views about how things are. You might be fully opinionated about how things are and yet be unsettled about what to do. Your views about what to do are reflected in the hyperplans your plan-laden belief state rules in or out.

How exactly? Let us bring in a term of art of Gibbard's, "the thing to do," which he stipulates to be expressive of planning states. Take

(4) Holmes thinks packing is the thing to do.

Following Gibbard, we're understanding this to mean that Holmes is resolved on packing, that he plans to pack.<sup>6</sup> We can get a handle on how hyperplans are put to work by asking: what does Gibbard say are the truth-conditions of (4), stated in terms of his model of plan-laden belief?

It is useful to first mention a wrong answer to this question. One might think Gibbard's idea is that (4) is true iff

For all  $\langle c, h \rangle \in P$ :  $h$  permits only options that entail packing

<sup>6</sup> Why then don't we just talk about the ordinary sentence 'Holmes plans to pack' instead of (4)? I take it Gibbard fixates on (4) because he is setting up to eventually talk about something like (3), which resembles (4) more closely. More on the relation between (3) and (4) below.

On this reading, Holmes's thinking packing is the thing to do is a matter only of what hyperplans are left open by the fact-plan alternatives in  $P$ . Thus, we have sort of the formal opposite of the case of straight factual belief: here it is the factual, centered worlds dimension that goes idle. Gibbard is sometimes described as a "pure" expressivist, where this is meant to contrast with the idea of a mixed or "hybrid" expressivism, which would hold that to be in a normative frame of mind is to be in a state that mixes the cognitive (doxastic) with something noncognitive (nondoxastic).<sup>7</sup> If one viewed Gibbard through this lens, an interpretation of his formalism like this might seem natural.

But this is not Gibbard's idea. Actually, this idea doesn't really make sense, because it doesn't really make sense to say a hyperplan permits (or requires) something *simpliciter*; rather, it does so only relative to a choice of a centered world. Instead, Gibbard's idea is that (4) is true if

For all  $\langle c, h \rangle \in P$ :  $h(c)$  permits only the options  $o$  in  $O_c$  that  
entail packing by the agent that  $o$  is centered on

The idea is to go to each fact-plan possibility compatible with Holmes's state and ask whether, when you evaluate the hyperplan component of that fact-plan possibility at its centered world component, the output of the hyperplan permits only packing outcomes. If so, then (4) is true: Holmes thinks packing is the thing to do. In this way, thinking that packing is the thing to do is a complex property of the individual fact-plan possibilities one's state leaves open.

In one way, there is a similarity between the state of thinking that something is the thing to do and the state of wanting, as we modeled it earlier following Stalnaker. Again, when you say somebody wants something, you are effectively saying that the pair of their doxastic state and their preference ordering satisfies a certain complex property; you are not just saying something about their preference ordering. In a somewhat similar way, for Gibbard, if you say somebody thinks something is the thing to do, you are effectively ascribing a complex property to the combination of their factual beliefs and their plans, not just saying something about the hyperplans they rule in or out. The truth of (4) turns partly on Holmes's prosaically factual beliefs. (To this extent Gibbard is already an expressivist of the hybrid style.<sup>8</sup>) On reflection, this should be obvious: one cannot think that packing is the thing to do unless one takes oneself to have the option to pack, and whether one has that option is a factual matter.

In another way, there is a formal disanalogy between the state of thinking that something is the thing to do and the state of wanting. Whether an agent is in the former kind of state is

<sup>7</sup> See, for instance, Schroeder (2009).

<sup>8</sup> There is of course still an important difference between Gibbard's expressivism and the hybrid-style realist-expressivism of, for instance, Copp (2001). The latter goes in for a realist metaphysic of normative properties, whereas Gibbard does not.



a matter of whether each of the fact-plan alternatives their state of mind leaves open, taken individually, satisfies the relevant condition. By contrast, there is no set of “desire alternatives” such that what the agent desires is a matter of what holds throughout those alternatives.

Let me stress a key difference between Lewis’s kind of enrichment to logical space and Gibbard’s. Lewis paired possible worlds with centers to capture belief *de se*. Belief *de se* is still factual belief, in the sense that there is a fact of the matter as to what centered world(s) you actually occupy. Your belief state is doing well when it doesn’t exclude centered worlds centered on you. When Gibbard pairs centered worlds with hyperplans, though, the resulting space is not fully factual, in the sense that there is no fact of the matter as to which fact-plan alternative you occupy, owing to the hyperplan component. There is, let’s presume, an actual world, and there is a fact of the matter about how you are situated in that world, but there is no “actual hyperplan.”<sup>9</sup> There is no fact of what to do, such that your plan-laden state of belief is doing well if it characterizes that plan. This gets at the point that the role of the plan-laden part of your belief state isn’t to represent plan facts. Again, your view about what to do is not a view about how things are.

#### 1.4. Planning and Normative Thinking

What has all this modeling of planning to do with normative thinking? Let’s put these three sentences again in front of us:

- (3) Holmes thinks he ought to pack.
- (4) Holmes thinks packing is the thing to do.
- (5) Holmes plans to pack.

It seems it should come out that (4) and (5) have basically the same truth-conditions, because of how Gibbard stipulates the meaning of “thing to do”; and we just said what those truth-conditions are. The tour through planning was supposed to result in a story about normative judgment, the sort of state ascribed by (3). Is Gibbard proposing that (3) has essentially the same truth-conditions as (4)/(5)? Is Gibbard’s model of this normative state just the state we have identified for (4)/(5) using hyperplans?

Gibbard is certainly angling for the idea that “we can understand normative judgments as an aspect of planning” (Gibbard 2006, 732). He says:

If I think that something is now the thing to do, then I do it. My hypothesis about ordinary *ought* judgments is that they are judgments of what to do, of what is the thing to do. I don’t, then, think that I ought right now to defy the bully unless I do defy him. If I fail to defy him, then as a matter of the very concept of *ought*, I don’t believe I ought to.

(Gibbard 2003, 153)

<sup>9</sup> Thus, I called the hyperplan component a “nonfactual parameter” in Yalcin (2011).



There are three concerns one might have about tying planning and normative judgment so closely together. Going through these will help to bring the picture out.

#### 1.4.1. Problem: Third-Personal Oughts and Their Connections to Plans

The first concerns third-personal oughts and their connections to plans. Suppose Holmes thinks packing is the thing *for Watson* to do, but not for himself. How does Holmes's judgment here about the thing to do tie into Holmes's own planning, given that he obviously can't literally decide for Watson?

Gibbard replies that in the relevant sense, Holmes can decide for Watson—he can decide what to do when in Watson's position. Roughly I understand Gibbard to model as follows: let's say *WATSON* is a function that takes a centered world and shifts its center to Watson (so it maps  $\langle x, w \rangle$  to  $\langle \text{Watson}, w \rangle$ , for any  $x$ , at least when Watson can be found at all in  $w$ ). Then Holmes thinks packing is the thing *for Watson* to do just in case

For all  $\langle c, h \rangle \in P$ :  $h(\text{WATSON}(c))$  permits only the options  $o$  in  $O_{\text{WATSON}(c)}$  that entail packing by the agent  $o$  is centered on<sup>10</sup>

That gets at Holmes's view about what to do when in Watson's shoes (as Holmes takes those shoes to be). We don't routinely speak of deciding and planning for others, and insofar as we do, it doesn't come as freely as our talk of the things others ought to do. I understand Gibbard to grant this. He will say that he means to get at the underlying structure of planning thinking. The link between this structure and our language might be indirect, just as the link between preference orderings and statements of want is indirect.

#### 1.4.2. Problem: Failing to Plan to Do What You Think You Ought

Second problem: can't you think you ought to do something without planning to do it? Gibbard explains this by saying that we are sometimes fragmented in our normative thinking:

A person often isn't "of one mind" in accepting a plan or not. For a crucial sense of 'ought', I say, the following holds: if you do accept, in every relevant aspect of your mind, that you ought right now to defy the bully, then, you will do it if you can. For if you can do it and don't, then some aspect of your mind accounts for your not doing it—and so you don't now plan with every aspect of your mind to do it right now. Whatever aspect of your motivational system issued in your doing otherwise didn't accept the plan to defy him right now. And so, it seems to me, there's a part of you that doesn't really think you ought to. You are of more than one mind on whether you ought to defy him.

(153)

---

<sup>10</sup> A variant of this would replace *WATSON* with something like an individual concept reflecting Holmes's mode of presentation of Watson.

This seems to trace the problem to the idealization of the model, rendering it perhaps analogous to the sort of idealization we were already involved in when modeling a doxastic state with a set of possibilities. The problem of logical omniscience afflicting our model of belief might also owe to the fact that realistic agents can be of “more than one mind” about something—fragmented and compartmentalized, as Stalnaker (1984) and Lewis (1988) discuss. Whether this is a big problem for the model depends on how much of it survives when we complicate it enough to reduce or eliminate the idealization. The kind of fix Stalnaker and Lewis envisage on the doxastic side seems to preserve intact the key idea of modeling with “a way things are according to the agent” (though it’s certainly a matter of debate); perhaps Gibbard can make a response in a similar shape, representing a divided mind with distinct sets of fact-plan alternatives. For the purposes of this paper, I treat this objection as contained.

#### 1.4.3. Problem: Deciding among Several Permissible Options

The third problem is the one that will take up the most space. Can’t you plan to do something without thinking you ought to do it? This worry is to do with the fact that plans frequently involve arbitrary choices. Suppose I plan to sit in the couch over there, by walking to it. I execute this scheme, deciding to take the first step with my left foot. I decided to step left first, but the choice was arbitrary; in my view, Righty was equally up to the task.<sup>11</sup> Had I stepped right instead, the world would not have been lesser to me in any way. My aim was just to get to the couch, and one cannot get to the couch except by some route. The thing is that while I decided left foot first, it’s not that I thought I ought to step left first. Where is the space in Gibbard’s model for this difference?

To be clear, the objection here isn’t that Gibbard can’t model the difference between thinking an option required and thinking of it as one permitted option among several. His hyperplan apparatus is perfectly designed to model that difference—remember that a hyperplan may permit more than one option.<sup>12</sup> Rather, the objection, flatfootedly, is that what you actually plan to do might be more specific—might be strictly more resolved—than your view about what you ought to do, and therefore your planning state and your state of *ought-thoughts* are not really the same, and therefore the truth-condition we sketched for (4) won’t carry over directly to (3). And therefore we remain in the lurch about (3), when really that was what we were out to analyze in the first place.

We could also put the problem using a piece of terminology from Gibbard (2006). Let’s say a *strategy* is a hyperplan that permits exactly one option for each contingency. Then the point is that, at least restricted to the occasion I face, I have in some clear sense both a strategy and a plan: my strategy calls for left foot first, whereas my plan—the part of me that

<sup>11</sup> We can say I thought about it for a second, if that makes it easier to describe me as “deciding.” The present issue doesn’t turn on which decisions or paths of action are selected “subpersonally.”

<sup>12</sup> His model also has no problem modeling the difference between lacking a view about which of several alternatives is permitted and thinking them all permitted.

reflects what I view as permitted and required—just says to step with some foot or other, permitting either foot. If we grant that I decided to step left, then it seems that the part of me that embraces a plan permitting multiple alternatives is not identical to my decisional state.

This draws the ambiguity in “thinking what to do” out somewhat. Explaining the state of *thinking what ought to be so* in terms of the state of *thinking what to do* can seem like a way of analyzing a state specified using a normative concept in terms of state that is not so specified. But while *thinking what to do* can mean *deciding what action to perform*, it can also just mean *thinking what is to be done*, with the infinitive conveying an implicit normative modality. In other words, “thinking what to do” can just mean “thinking what ought to be done”—in which case the former marks no special progress. (“The house is to be cleaned” can be heard as a prediction, but when said to the housekeeper, it rings as normative. Gibbard’s discussion of “to be desired” (Gibbard 2003, 22) displays sensitivity to just this kind of contrast.) There evidently are these two different things “thinking what to do” can mean. Obviously, we want to make sure that any theory that explains *thinking what ought to be the case* in terms of *deciding what action to perform* doesn’t rely along the way on this ambiguity.

Gibbard says some things that seem to speak to this issue.<sup>13</sup> He agrees that there is this difference we are talking about:

Is there a difference, then, between rejecting an alternative—not permitting it to myself—and just not choosing it? Surely there is. The two differ in “valence” or oomph. To think this distinction intelligible, must I already think that permitting myself an alternative consists in attributing some special kind of property to it? No, distinguishing in this way is clearly a part of planning, but there is no need to think, at the start of inquiry, that distinguishing this way is a matter just of factual belief. My claims here concern what one commits oneself to in planning, and the facts I’m allowing at the outset are straightforward and prosaic. We can distinguish preference and indifference without first admitting facts of a kind more ethereal.

(Gibbard 2003, 153)

Put aside the issue of whether a normative realism would work better for explaining the difference we’re getting at—our interest is just in understanding Gibbard’s alternative to an explanation based on “ethereal” facts. He indicates here that the difference is one tied up with the difference between preference and indifference, a difference he stresses earlier in the book:

Buridan’s ass might have been wiser, and a wiser ass would choose one bale of hay or choose the other. She wouldn’t thereby rule out choosing either—or at least there’s an important sense in which she wouldn’t. She wouldn’t be in disagreement with plumping for the other from indifference. It is in the nature of planning, after all, to distinguish rejecting an alternative

<sup>13</sup> I am especially indebted here to conversations with Sophie Dandelet, and also to her Dandelet (2017).

by preference from simply not choosing it in that, from indifference, one chooses another. Rejecting an alternative is something more than just taking a different alternative when there is more than one alternative that one doesn't reject by preference.

(Gibbard 2003, 55)

The idea is that when I stepped left first, I did it out of indifference. In a counterfactual scenario where I acted the same but permitted myself only to step left, Gibbard seems to say that I do this by preference. In these two scenarios, I act the same but I am representable by different sets of fact-plan alternatives, and ultimately this difference is grounded, Gibbard appears to say, in my state of preference.

I pause exegesis mode here to say that this injection of preference into the story surprises me. Initially it seemed the proposal was going to be that planning, understood as a distinct state of mind from preference, is what normative judgment is tied up with. But Gibbard seems to say that what one views as permitted is in some sense explained by what one prefers, since he seems to be saying that an agent's viewing several options in a situation as permissible entails that the agent is preferentially indifferent among the options. That's a very strong tie to hypothesize between planning and preference. (I won't dwell on Gibbard's motivations here, which seem to do with how he hopes to account for when agents with different views about what to do count as disagreeing.)

Be all that as it may, this detour through preference doesn't address the problem we started with. Again, I decided left foot first, but it's not that I thought I ought to step left first. I planned to step left first, but it's not that I thought stepping left was *the* thing to do. Even granting that my stepping was out of indifference, the problem is that we still have a psychological gap between the part of me that has (indifferently) settled how I act—I decided, I picked a strategy—and another part of me that reflects what I think would have been permissible. The problem isn't "how can we make sense of me embracing a plan that permits several options, and distinguish that from embracing just one." The problem is rather that it can be that I both (1) view several options as permitted and also (2) have decided which to select. One set of fact-plan alternatives cannot reflect these two facts about me, then. From a modeling point of view, we need a set of fact-plan alternatives to handle the state of me that corresponds to (1), and another set for the state of me that corresponds to (2). Evidently there's my state of normative judgment, which pronounces on what's permissible, and there's my decisional state, which seems more directly related to what exactly I end up doing. Maybe these two states have a similar logical structure, both to be elucidated with hyperplans, and maybe they tie into each other in intimate ways—for instance, perhaps (ideally) you only decide to do what you take to be permitted—but it is hard to see how they could shake out to be the same state.<sup>14</sup>

14 Discussing Gibbard (1990), Railton sees what is maybe a similar gap: "If there is an element of language that is purely action-guiding, I suspect it is closer to 'the thing to do' than to 'the rational thing to do', or to 'the

Once we grant the gap here, we can flag two ways that decision and normative judgment might intertwine. First, a normative view might itself be the result of decision. (Maybe “deliberation” is the word we’d more naturally use.) We speak of deciding, not just what to do but also what is permitted and required (of us, or of others). The result of a decision might be an intention to act, or it might be a normative judgment to the effect that a certain action is called for. When we talk about “deciding what to do,” we perhaps usually mean to be talking about a state that results in an intention to act in some way. But we could also mean “deciding what is to be done,” where the result of that is foremost a normative judgment, perhaps by an agent not yet resolved on a particular course of action. Second, if we think that one can only decide to do what one takes to be permitted (a view it seems Gibbard would find attractive, at least for agents in some sense ideal), then although deciding and normative judgment are strictly different things, a decision to act always goes with a normative judgment to the effect that the act is permissible.

Reconsider now:

- (3) Holmes thinks he ought to pack.
- (4) Holmes thinks packing is the thing to do.
- (5) Holmes plans to pack.

Again, take it we have these truth-conditions for (4):

For all  $\langle c, h \rangle \in P$ :  $h(c)$  permits only the options  $o$  in  $O_c$  that entail packing by the agent  $o$  is centered on

We were asking whether these truth-conditions are basically right, according to Gibbard, for (3).

If what we said in recent paragraphs about the gap between planning and normative judgment is on track, then the state ascribed in (5) is not the same as the state ascribed in (3), though they might be importantly connected. Now when Gibbard introduces “the thing to do” talk, he stipulates that it is to be expressive of decisions.<sup>15</sup> That stipulation

---

thing it makes most sense to do” (Railton 1992, 966). See also Scanlon (2006). Engaging Railton, Gibbard seems to grant the possibility of a contrast between *being the thing to do* and *being the rational thing to do* (Gibbard 2003, 152), which sounds like the contrast between decision and normative judgment that I am asking about.

- 15 He writes: “Suppose, let me stipulate, the phrase works like this: to conclude, say, that fleeing the building is *the thing to do* just is to conclude what to do, to settle on fleeing the building. By sheer stipulation, then, the meaning of this phrase ‘the thing to do’ is explained expressivistically: if I assert ‘Fleeing is the thing to do,’ I thereby express a state of mind, deciding to flee. I then proceed to ask how language like this *would* work. In the back of my mind, of course, is the hypothesis that important parts of our actual language do work this way. Mostly, though, I don’t argue for this hypothesis; rather I ask whether the hypothesis is coherent

pushes us to see (4) and (5) as equivalent, and therefore it pushes us to see (4) and (3) as inequivalent—which would recommend the position that the truth-conditions of (4) are not the same as the truth-conditions for (3).

What's a charitable interpreter to say here? It certainly seems that Gibbard's proposal is at least that the truth-conditions of (3) *mimic* the ones we have stated for (4). Holmes is in a (hyper)plan-laden state of belief that reflects his normative judgments, his views about what is permitted and required in various situations. Model this with a set of fact-plan pairs  $N$ , which we hold in our minds for the moment separately from  $P$ . Then, roughly, (3) is true just in case:

For all  $\langle c, h \rangle \in N$ :  $h(c)$  permits only the options  $o$  in  $O_c$  that entail packing by the agent  $o$  is centered on

Is the normative state modeled with  $N$  exactly the planning state modeled with  $P$ , the related state of Holmes that we characterize when we say something like (4)? For the reasons reviewed, it seems we shouldn't say that, but as a textual matter, I am not sure how to read Gibbard on the question. Certainly, he wants to say that his notion of planning is not exactly the ordinary notion, for the reasons we reviewed in the previous two sections: it's an ideal model aimed at capturing the abstract structure of contingency planning. The point we're running up against right now, though, grants this, and still observes that there is a gap between this ideally conceived planning and ideally conceived normative judgment. Does Gibbard grant this point? If he does, then I don't know where he makes it with a reassuring level of emphasis. You've seen the quotes above tying planning and *oughts* together, which pull the other way. Consider however this passage:

Reserve the term 'ought' as a quick way of saying "has most reason." The unqualified dictum I started with was this: to believe that one ought to do a thing—that one has most reason to do it—is to decide to do it. This Scanlon rejects, and rightly; it needs the qualifications I have just been stating, and which I stated in the book. My slogan to a closer and more verbose approximation might be this: to believe that a person ought to do a thing is to require it of oneself for the hypothetical case of forthwith being in that person's precise situation.

(Gibbard 2006, 731–32)

The unqualified dictum explains a state of normative judgment in terms of a decisional state not specified with the help of normative vocabulary (viz., deciding to do a thing), whereas the more precise statement explains the state using normative vocabulary (invoking talk about what one "requires" of oneself). Maybe this favors the reading that Gibbard would

---

and what its upshots would be. Only much later in the book do I turn to our actual everyday thoughts and ask if the shoe fits" (Gibbard 2003, 8).

say that (3) doesn't shake out to be quite the same as (5), though there is a deep affinity in underlying structure.<sup>16</sup>

To put the question again: is the normative state modeled with  $N$  the same as the planning state modeled with  $P$ , the related state of Holmes that we characterize when we say something like (4)? Since I want Plan A to align with Gibbard, and since I seem to find conflicting tendencies in what he says, my official stipulation will be that Plan A is agnostic on this crucial question.

Anyway, the thrust of it is this. Belief states are plan-laden in at least the sense that they have a structure to be articulated with hyperplans as above: they are states that cut the space of fact-plan alternatives. One's normative judgments are explained essentially in terms of this plan-laden structure—specifically, along the lines of the truth-conditions recently set down for (3). The not-fully factual character of these judgments traces to the conditions they place on the fact-plan alternatives they rule out, conditions partly a function of the (nonfactual) hyperplan dimension of those alternatives. We have here an alternative to the realist idea that states of normative judgment are states that represent normative properties or facts.

So we have, finally, a view on the table about what Holmes's state of mind is like when (3) is true, according to Gibbard.

### 1.5. Gibbard's Metatheory

Besides his novel formal model of states of normative judgment, Gibbard offers a distinctive kind of philosophical gloss on, or metatheory for, his model.<sup>17</sup> The gloss is interestingly different from the sort of gloss we find from other theorists in the tradition of decision-theoretic/possible worlds modeling, for instance Lewis or Stalnaker, who Gibbard otherwise looks to be rather continuous with. Those theorists would say that to believe that grass is green is to be in a state that rules out possible worlds wherein grass is not green, and they would hold that these possible worlds that the state rules in or out are, at least eventually, explicable independently of intentional mental states—they are not themselves explained in mental terms, not themselves metaphysically dependent on anything mental (Lewis 1994; Stalnaker 1984). Gibbard is skeptical, however, that a nonintentional notion of modality is available for this purpose (see, for instance, Gibbard 2012, 277). He prefers not to attempt reduction. He explains the “possibilities” that mental states rule out *as themselves mental states*. He has this view quite apart from his proposal to model in terms of hyperplans—this is how he would want to think about an ordinary, hyperplan-free possible worlds model of belief. Gibbard will agree that to think grass is green is to “rule out a possibility,” but fundamentally he will explain this state as the state of ruling out another mental state, the state of rejecting grass is green—the

<sup>16</sup> The “qualifications” Gibbard alludes here to have to do with the first two problems we discussed, though, so we shouldn't read this quote as directly animated by the problem we're focused on in this section. So I don't put too much weight on this passage.

<sup>17</sup> In this section, I draw on some of the ways I put things in Yalcin (2018).



mental state of *rejecting that grass is green* is the “possibility” ruled out. Similarly, he would describe the state of believing that grass is not green as “disagreeing with believing” that grass is green—as rejecting believing grass is green. The centered worlds of the model are interpreted by Gibbard as maximally opinionated states of (factual) belief—they are not, as Lewis or Stalnaker would have it, maximally specific ways things might have been, understanding the relevant modality as fundamentally nonmental. Likewise, he glosses the hyperplans of his model as maximal states of decision—the states that idealized “hyperplanners” would be in.

This will seem to some like a dangerously tight circle: it is a model of mental states whose basic resources for modeling mental states include mental states. Out is the idea of characterizing propositional attitudes as relations to contents, if the latter are understood in a traditional way as some sort of mind-independent abstracta—sets of possibilities, for instance, as Lewis and Stalnaker would have it. We don’t arrive on this picture at a conception of mental content giving us a handle on it in other terms; understanding must somehow come from the whole system. Schroeder (2008a) argues that all this renders Gibbard’s view explanatorily deficient. Does Gibbard’s view leave it mysterious what makes it the case, when it is the case, that one content is incompatible with another? For example, the state *believing grass is green* and the state *believing grass isn’t green* “disagree” with each other; they are in logical tension. In virtue of what? Not, says Gibbard, in virtue of their having incompatible contents. What, then? Gibbard says he has no further explanation of such disagreement facts; he takes them as primitive.

Gibbard argues that this is not a disadvantage, however, because the orthodox line of explanation invokes “substantial, unexplained truth, eschewing any minimalist explanation of truth” (Gibbard 2003, 74). He expands:

Proceeding this way might seem to be philosophical theft. The scheme amounts just to helping ourselves to the notion of disagreeing with a piece of content, be it a plan or a belief. A negation, we say, is what one accepts when one disagrees—and this explains negation. Now I wish, of course, that I could offer a deeper explanation of disagreement and negation. Expressivists like me, though, are not alone in such a plight. Orthodoxy starts with substantial, unexplained truth, eschewing any minimalist explanation of truth. I start with agreeing and disagreeing with pieces of content, some of which are plans. It’s a thieving world, and I’m no worse than the others.

(74)

In his more recent book (Gibbard 2012), he calls orthodoxy “Fregeanism,” and puts the problem, or at any rate one key problem, for the view like this:

Not all impossibilities make for entailment—or at least they don’t make for the kind of entailment that, with enough conceptual competence, a thinker can recognize. It is this kind of entailment, inconsistency, and the like, we might well think, that the Fregean needs to explain, or the whole Fregean project fails. The thought I’M DRINKING WATER doesn’t recognizably entail I’M DRINKING H<sub>2</sub>O, unless one knows the chemical composition of water. The

thought I'M USING THIS LECTERN likewise needn't recognizably entail I'M USING AN ORIGINALLY WOODEN LECTERN, even if it couldn't be that this very lectern was originally made of anything but wood. The metaphysical impossibility of a putative state of affairs needn't yield its conceptual impossibility.

(277–78)

The hypothetical maximally opinionated agents at the bedrock of Gibbard's theory aren't explained in nonintentional terms. But Gibbard says orthodoxy requires a logical space for marking epistemic or conceptual distinctions that go beyond metaphysically possible distinctions—a space where the possibilities must be intentionally specified. So, he thinks we have at worst parity between the views here.

There is one more component to Gibbard's meta-theoretical reflections. He styles his preferred philosophical gloss on his model as part and parcel of an expressivist approach. He says things like this:

The orthodox explain disagreeing with a claim as accepting its negation, whereas I go the other way around: I explain accepting the negation as disagreeing with the claim. Agreement and disagreement are what must ground an expressivistic account of logic.

(Gibbard 2006, 73)

One gets the impression that to go expressivist in the relevant way about normative judgment, it is not enough to embrace a model of this state of mind that renders this thinking as not (just) a matter of the way one represents the world to be. It is not enough to model with fact-plan alternatives, tracing the not-fully factual character of normative language at the modeling level to their interactions with hyperplan structure. One must also adopt a certain foundational gloss on this model—an “attitudes-first” metatheory like that just described. The fact-plan possibilities of the model must be understood intentionally, as idealized “hyperdecided” states of mind; and the relations of (dis)agreement between them and between less-than-maximally decided states are to be taken as primitive.

So the picture of normative judgment I've just sketched is Plan A. Most prominently Plan A provides a formal model of normative states of mind and a certain kind of philosophical gloss on that model. I mean Plan A to be a highly selective take on Gibbard's picture of normative judgment, with an emphasis on the role of hyperplans in the story that, I should stress, is quite out of proportion to Gibbard's own emphasis. I have skirted over many nuances.<sup>18</sup> But if I haven't got Gibbard's view just right, I hope you will agree that Plan A is very Gibbardian, and worth assessing.

<sup>18</sup> At several turns, especially in discussing “facts” and “the way the world is,” I have ignored qualifications whispered to me by the quasi-realist devil on my shoulder, reminding myself that Gibbard's talk of *prosaically factual* belief is suggestive of the sort of grip on Reality my discussion seems to presuppose.

## 2. Plan B

Now to Plan B. Plan B is like Plan A in explaining normative judgment as plan-laden—that is, modellable with the help of things like hyperplans—and as not in the business of representing normative facts. It is an expressivist view. But it differs from Plan A at the foundational level in two key ways. First, Plan B embraces what Gibbard styles as the “orthodox” or “Fregean” metatheoretic attitude toward the formal model. Second, Plan B explicitly says that normative judgment isn’t planning, understood as deciding how one will act, though it allows that normative judgment may be formally analogous to planning.

Let me now go through these differences. I don’t attempt any full-throated defense of Plan B against Plan A here. My real aim is just to bring out Plan B as an option, so we have a sense of which choices on the expressivist road are separable and which come together.

### 2.1. *Orthodox Metatheory*

It is one thing to model as Gibbard does, and another thing to interpret the model as Gibbard does. You can model as Gibbard does, without interpreting the model as Gibbard does. It would be a mistake to reject Gibbard’s model merely because one isn’t ready yet for his distinctive metatheory. We have to separate these things.

Plan B says that when we are thinking about the ingredients of our formal model, we don’t understand possible worlds, centered worlds, hyperplans, or fact-plan possibilities as themselves mental states; we don’t construe the model as explaining the content of a mental state by reference to further contentful mental states; we don’t treat it as a brute fact that the state of believing  $p$  is in logical tension with the state of believing  $\neg p$ . Possible worlds, we take it, are not maximally opinionated mental states, though we might use a possible world to model a state of maximum opinionation. Plan B says: go back to what we might think of as the more typical way of understanding possible worlds modeling, familiar from (e.g.) Lewis (1979, 1994) and Stalnaker (1984)—an approach Gibbard would group with the “orthodoxy.”

As noted above, Gibbard gives some indication that he thinks that this approach really smuggles something it claims to explain through the back door—the truth-conditions of a sentence must be said to discriminate between “conceptual” possibilities, possibilities that are intentionally specified (Gibbard 2012, 277–8). This is part of why he holds his view to be not less explanatory than orthodoxy. If this is at the crux of the debate, I wish Gibbard had given the topic more airtime. While I doubt that the orthodox view does require an intentional conception of possibility—there are certainly various well-known ways to ensure, for example, that I’M USING THIS LECTERN doesn’t entail I’M USING AN ORIGINALLY WOODEN LECTERN, and so forth, without such a conception—this is a suboptimal place to debate that large issue.<sup>19</sup>

<sup>19</sup> I objected to Gibbard’s preferred metatheory in Yalcin (2018), but I now think I didn’t adequately address the appendices of Gibbard (2012), in particular the worry about whether the logical space assumed by orthodoxy can really be nonintentionally characterized.

Gibbard is playing defense, not offense. I'm satisfied to observe that the orthodox position is not subject to a new style of objection, one that Gibbard's own view is somehow immune to. Either the score is tied—both Plan A and Plan B both traffic in partly in unexplained intentional notions—or Gibbard is wrong and Plan B can be carried out with a nonintentional notion of logical space. Since I am only plumping for the view that Plan B deserves airtime along with Plan A, I can live with a tie for the purposes of this paper.

There are many questions one could raise about what it is to “explain” or “ground” an abstract model. One basic sort of question we could ask about a simple possible worlds model of belief is this:

In virtue of what sort of facts is an agent in a state of belief well-modeled with such-and-such possible worlds content rather than some other possible worlds content?

It seems worth saying that this sort of question has hardly gone unaddressed by those who enjoy the orthodoxy style of model-interpreting. The basic issue here was the stuff of much of the philosophy of language and mind of the eighties and nineties on the metaphysics of content. Restricting attention to the broadly decision-theoretic modeling tradition Gibbard departs from, Stalnaker (1984) and Lewis (1994) both directly address this question about how facts of content should be understood to metaphysically depend or reduce to facts of another stripe. Their views differ in important ways, and resist easy summary. But the point I'm stressing is just that there is not a blithe indifference, or a failure to address the question, on the side Gibbard styles as orthodox. Instead, there's a nontrivial literature. To say that “Orthodoxy starts with substantial, unexplained truth” in the face of this work seems a touch glib—though maybe I'm missing what sense of “explanation” is intended.

There's a whole debate to have about this—elsewhere. For now, let me just repeat the point that if we truly are in the thieving world Gibbard says we're in, then Plan B and Plan A are at worst on par in the relevant respect, and therefore Plan B remains in the running as a view worth talking about. While I don't think Team Plan B should concede the antecedent of this conditional, I won't argue it now.

But can one embrace an orthodox attitude at the metatheoretic level and still be expressivist? Isn't Gibbard's style of metatheory constitutive of expressivism? Isn't “explaining in terms of mental states” what it is to be expressivist?

To repeat something I tried to say in Yalcin (2018), I think it is a mistake to take Gibbard's distinctive kind a metatheory to be the hallmark of an expressivist approach. I suggest we separate two ways of understanding how an expressivist might be described as “explaining in terms of mental states.” One way is in the vein of Gibbard's metatheoretic reflections: on this way we “explain” the “possibilities” the content of a mental state eliminates as themselves mental states, and take (dis)agreement between mental states as primitive. But there is second way of understanding how an expressivist view might bring mental states into

explanation. This is the idea that expressivism is characterized by a strategy we could call *psychological ascent*. Here is a crude way to put the recipe:

**Expressivism by psychological ascent.** To go expressivist about  $\varphi$ , first reject the question “What is the world like when  $\varphi$  is the case?” Replace it with the question: “What is the state of mind of accepting  $\varphi$  like?” Answer this question in such a way that the state of mind is understood as not tantamount to ordinary factual belief that something is the case—as not representing  $\varphi$ -facts. Then approach the target discourse from this perspective: find a way to elucidate the semantics and pragmatics of  $\varphi$  consistent with the idea that accepting  $\varphi$  is being in this not-fully-factual state of mind.

Psychological ascent is a pathway for stating an expressivist view. When I look at the various twentieth-century works in the expressivist genre, it looks to me if anything like the standard pathway. One can follow this path without also taking on Gibbard’s style of metatheory.

The expressivism about epistemic modality I have defended elsewhere (Yalcin 2007, 2011) is expressivist in this sense—as is Gibbard’s classic work on indicative conditionals (Gibbard 1981) when viewed at a natural angle. It is famously difficult to say what the world has to be like to make an indicative conditional true, or to make it true that something (epistemically) might be the case. The expressivist mode is to table that question and ask instead about the state of mind that goes with accepting what the sentences say. It says: don’t start with questions like this:

What is the world like when ‘It might be raining’ is true?

What is the world like when ‘If the marble isn’t under cup A, it’s under cup B’ is true?

But instead with questions like this:

What is to think it might be raining?

What is to think that if the marble isn’t under cup A, it’s under cup B?

Or, semantically ascending, with questions about attitude ascriptions, like this:

What is the world like when ‘A thinks it might be raining’ is true?

What is the world like when ‘A thinks that if the marble isn’t under cup X, it’s under cup Y’ is true?

This kind of expressivist ends up with a conception of the truth-conditions of the target attitude ascriptions that does not resolve them into relations to propositions characterizing the world as being some way or other. (For instance, they might say that when *A* *thinks it might be raining*, that’s because *A* is in a state of mind leaving rain possibilities open,

and not because *A* believes-true some *might*-proposition; or they might say that that when *A* thinks that the marble is under cup *Y* if it's not under *X*, that's because *A* has high credence in *Y* conditional on not-*X*, and not because there is some conditional proposition *A* believes.) Thereby this expressivist dissolves the original questions, the ones in search of facts to be expressed by the original epistemic modal sentences.<sup>20</sup>

Plan B takes this kind of approach to normative language. Don't ask:

What is the world like when 'Holmes ought to pack' is true?

Instead ask

What is it to think Holmes ought to pack? Semantically ascending: what is the world like when 'Holmes thinks that he ought to pack' is true?

Plan B, like Plan A, gives an answer that explains the ascription as not ascribing to Holmes a prosaically factual belief. In particular, it's not that Holmes locates himself in a world that includes the normative fact that he ought to pack. Rather, it's for Holmes's state of mind to be plan-laden in the right sort of way.

Obviously, "psychological ascent" is meant to remind you of "semantic ascent." They have a lot in common. Semantic ascent can be helpful in metaphysics. When two sides disagree about what some aspect of reality is like, that will often make for a difference between the two sides in respect of what they can say their terms represent. That in turn can make it hard to state the point at issue between the sides in a neutral way. Retreat to talk of sentences and what those sentences say can be a way of preventing the sides of the debate from talking past each other. But stereotypical applications of semantic ascent still do involve the assumption that the target sentences (the sentences which are the locus of ascension) have truth-conditions, that they say something or other about reality. The two sides might disagree about what is represented by sentences containing the terms key to their dispute—semantic ascent is just what will make that perspicuous—but in the standard examples, the two sides at least agree that the target sentences are in the business of describing the world as some way or other. Take for instance the "ontological debate" about whether Pegasus exists (Quine 1948). The Pegasus believer explains the meaning of 'Pegasus' in terms of a certain winged horse of their ontology, suppose. The Pegasus denier, lacking that thing in their ontology, says something else. We'd like to say that the two sides disagree about whether Pegasus exists, but annoyingly, it seems the sides must understand the meaning of a sentence like 'Pegasus doesn't exist' in different ways, owing to the different ways they will each explain the meaning of 'Pegasus'.

<sup>20</sup> Though, of course, the expressivist faces new questions in semantics and pragmatics, about how to think systematically about the communicative roles of these sentences in the absence of factuality. I discuss this especially in Yalcin (2012, 2018).

Retreating to talking about the conditions under which ‘Pegasus doesn’t exist’ is true can help to isolate the metaphysical dimension of the debate from the semantic dimension. But note that in standard cases like this, the two sides agree that the target sentences have truth-conditions; at worst they differ on what the truth-conditions are. Both the believer and the denier do think ‘Pegasus doesn’t exist’ characterizes the world as being some way or other, though they may differ in what way that is.

Psychological ascent is like semantic ascent in that we are retreating, broadly speaking, to talk of things with intentional properties (and which both sides in the debate are happy to recognize), but it is a move that is free of a presupposition that the target sentences are factual, in the sense of characterizing the world as being some way. The expressivist about normative discourse wants to reject questions like “What feature of reality corresponds to its being wrong to break promises?” The expressivist is not better served by the question we’d get by semantic ascent: “Under what conditions would ‘It is wrong to break promises’ be true?” The question “What is to think that it is wrong to break promises?” brings us to level of description where the expressivist can start to unfold their view.

The strategies of psychological ascent and semantic ascent can be fruitfully combined. The question we get by psychological ascent—“What is to think it’s wrong to break promises?”—is a metaphysical question as susceptible to semantic ascent as any other. Semantically ascending, we get the question “What is it for something like ‘A thinks it’s wrong to break promises’ to be true?” My Plan B expressivist favors this kind of two-step ascension. It is clarifying to get at the view this way, because as we’ve seen, our expressivist’s abstract model of normative states of mind does not map into attitude ascription in a trivial or linear way. You don’t just disquote to articulate the truth-conditions of attitude ascriptions; the right-hand side is substantive.<sup>21</sup> Mentioning rather than using the relevant attitude verbs prevents confusing slides between the modeling language and ordinary language. This perhaps reveals why I elected to place such emphasis, in the preceding sections, on framing Gibbard’s view as a position about the truth-conditions of attitude ascriptions that embed normative vocabulary.

I don’t mean to suggest that psychological ascent isn’t part of Gibbard’s own approach. He gives a theory of normative judgment, not normative facts, and the theory of normative judgment is supposed to dispel the need for normative facts; he makes the move of ascending to the psychological level ultimately to dispel philosophical perplexity about a seeming domain of facts. That embodies the core expressivist pattern, and the pattern my views about epistemic modality also fit. Mixed into Gibbard’s development of his theory is a separable body of metatheory, though. We do better to unmix these issues, I suggest, and consider the metatheory questions as different ones.

Sometimes expressivism is conceived of as something like a special kind of semantic theory, one that identifies the compositional semantic values of sentences with mental

<sup>21</sup> Gibbard understands his theory as combinable with a thoroughgoing minimalism about truth (e.g., Horwich 1998). Plan B, by contrast, is not so combinable.



states (Blackburn 1993; Rosen 1998; Schroeder 2008b; Charlow 2015; among others). One might understand it like this because one thinks that adopting the strategy of psychological ascent in the expressivist's way must lead, on its most plausible development, to such a semantic theory. If one has such a view, then one will be skeptical about unmixing the issues as I suggest. I argue against this view directly in Yalcin (2018). Less directly, I think the case of epistemic modality (Yalcin 2007, 2011, 2012) already illustrates that one might psychologically ascend in the expressivist's characteristic way without also signing up for a nonstandard kind of semantic theory, or a nonstandard way of interpreting standard semantic models.<sup>22</sup>

## 2.2. *Normative Judgment Not Reduced to Planning*

Plan B takes a stand on the thing Plan A was agnostic about. It agrees that planning and normative judgment might be alike in both calling for something like hyperplans in models of their content. But Plan B has no pretension to reduce normative judgment to planning. We earlier said, following Bratman, that planning is its own thing, not to be reduced to something else, like some combination of belief and desire. Now Plan B says the same thing about normative judgment. Normative judgment—or more specifically, one's views about what is permissible to do in various situations actual, hypothetical, and counterfactual—is its own thing, too, not to be reduced to something else, like some combination of belief, desire, and intention or planning. We can (and will) of course still attempt to *offer a model* of the state, and we can still aim to *offer an interpretation of the model* in an orthodox vein, as we might do for a decision-theoretic model of belief and desire by (for instance) clarifying the functional role of the state. It is just that we will theorize under the assumption that normative thinking is not identical to planning.

In this I seem to be on the same page as Scanlon (2006). Discussing Gibbard (2003), Scanlon writes:

The difficulties I have described do not arise from the expressivist strategy of giving a (non-reductive) psychological account of normative attitudes, but rather from the attempt to base this explanation on the single notion of a plan. My suggestion is that Gibbard's strategy could be more plausibly carried out if he were to broaden the range of notions that figure in his psychological explanation. These will include notions of an explicitly normative character, such as the idea of seeing something as a reason. But we can distinguish here, just as Gibbard proposes, between the normative content that these notions have when one employs them in deliberation and their descriptive employment in a psychological account of deliberating agents.

(726)

<sup>22</sup> I am inclined to read the (otherwise diverse) expressivist stylings of Stalnaker (2014), Santorio (2016), Starr (2016), Willer (2017), and Moss (2018) as compatriot views here.

I won't put the sort of normative state Plan B models as the state of "seeing things as reasons"—instead I want, again, to talk about one's views about what is permissible to do in various situations actual, hypothetical, and counterfactual—but the relevant point is that one can pick out the state using normative vocabulary without apology, and compatible with embracing an expressivist pattern of explanation of some target discourse in terms of a model of that state.

Would doing this leave something unexplained? It is good to compare the situation here to the situation with (prosaically factual) belief and preference. In giving an explanatory model of these states in the broadly decision-theoretic style, we don't have to claim to be reducing these states to states of other kinds in order for the modeling project to seem like progress. (One might have a reductive aim—to reduce everything to betting dispositions, for instance—but that is hardly a prerequisite for taking models of this sort seriously.) We limn the structure of these states with our abstract modeling tools and we say how the elements of the model are supposed to connect to each other and to other features of reality. One could argue about how illuminating the abstract models of agents pursued in things like decision theory are, but it would be strange to complain that the defect in these views is that they don't come with a recipe for reducing the states they traffic in to something else.

Perhaps unsurprisingly, I am not sure to what extent Gibbard would disagree with this part of Plan B. His "possibility proof" was primarily aimed at showing that the concepts needed to handle planning and disagreement in plan recommend a path for thinking about normative concepts; and one could succeed in showing this without reducing normative judgment to planning.<sup>23</sup> He's clearly saying that a formal entity approximating the structure of a hyperplan is of use for modeling both planning states and normative thinking, and on that Plan B agrees. Just as a textbook possible worlds theorist might hold that two distinct kinds of intentional states, such as belief and (say) imagination, can be helpfully modeled by appeal to a formal object of a single kind (a sets of possible worlds, say), so a theorist might hold that planning and normative judgment, while different kinds of mental states, both have at an abstract level a structure that is helpfully modeled using hyperplans.

In any case, the thing Plan B sticks its neck out on is normative thinking, not planning.

### 3. Plan B+

Come back yet again to this sentence:

- (3) Holmes thinks he ought to pack.

We said Gibbard models it like this: (3) is true iff

<sup>23</sup> "A fully consistent planner in my sense of the term, I tried to show, would in effect deploy concepts that work much as a non-naturalist would think that normative concepts work" (Gibbard 2006, 735).

For all  $\langle c, h \rangle$  left open by  $N$ :  $h(c)$  permits only the options  $o$  in  $O_c$  that entail packing by the agent  $o$  is centered on.

Again,  $N$  is a set of fact-plan possibilities determined by Holmes's state of mind at the evaluation world for (3).

Now, the most basic Plan B position embraces the ideas of the last section (pass on Gibbard's metatheory, embrace a metatheory of a more orthodox kind, psychologically ascend and then semantically ascend to state the position using a nondeflationary notion of truth, and take normative judgment to be its own thing) but keeps this specific modeling proposal for ought-thoughts—it keeps the truth-condition above. That is, this most basic version of Plan B disagrees with Plan A not on the modeling questions but on the philosophical gloss on the model. The two plans agree about how to model normative judgment with hyperplans.

But now that we've come this far, it is interesting to consider a version of Plan B that departs from Plan A also in some of the formal modeling respects. The idea of a hyperplan is a very interesting one, and it seems to me to open up some fruitful questions on the modeling side of things. Metatheoretic issues tend to get most of the airtime in discussion of Gibbard's approach, but the formal model he gives is itself worthy of investigation quite apart from those issues. Now I want to explore some modifications to Gibbard's modeling proposal, to do with how exactly hyperplans figure in the story. So, I will be exploring ways of giving truth-conditions to (3), which invoke hyperplans but not quite in the way Gibbard does. The ultimate formal model of normative judgment we will end up with is still very much in the vein of Gibbard's, but it will come apart from the letter of Gibbard's own version in key ways. If we add the modifications of I'm about to suggest to the ideas of the last two sections, let's say what we end up with is Plan B+.

If your main interests were on the metatheoretic side of things, this is a safe time to check out of the paper.

### 3.1. *Hyperplans as Information-Based*

Gibbard models an agent's state of belief-and-normative-judgment with a set of fact-plan possibilities. Question: why exactly are centered worlds and hyperplans packaged together in this way? Let's ask:

Why model an agent's state with a single set of fact-plan possibilities, rather than with two separate sets—a set of centered worlds (modeling ordinary factual belief) and a set of hyperplans (getting at the agent's normative views)?

To think about this, consider the following kind of example. Suppose that  $\langle c, h \rangle$  and  $\langle c', h' \rangle$  are the only two fact-plan possibilities compatible with Holmes's state of mind. Suppose also that Holmes thinks that he ought to pack. Now we could ask: what does  $h$  permit relative to  $c$ ? And what does  $h'$  permit relative to  $c$ ?

For all Gibbard says, the following could be the case:

- $h(c')$  forbids packing
- $h'(c)$  forbids packing

Suppose that is so. So Holmes thinks he ought to pack (this holds throughout his plan-laden state), and yet his state of mind leaves open plans that say not to pack relative to centered worlds that are bonafide doxastic alternatives for him.<sup>24</sup>

This not a contradiction, but it is a bit puzzling and counterintuitive. We might be able to make sense of this if the different centered worlds compatible with Holmes's state of mind potentially fixed different sets of options. However, it is hard to understand how they could, given the way Gibbard talks about options:

An occasion, as I have characterized it, contains much that the agent has no way of knowing, but one's plans must respond to features of the occasion available to the agent. Alternatives must be subjectively characterized, so that the same alternatives are available on subjectively equivalent occasions. And a plan must permit the same alternatives on subjectively equivalent occasions. (57)

"Keep away from radioactivity" would surely be a good part of a plan for living, if only we knew how to tell what's radioactive—but it's not much help if we don't. Plans, it seems, must be couched in terms of features that we can recognize: features of contingencies and features of options. Both of these must be available to the person who follows the plan. "Buy low, sell high" is no plan we can implement. Plans must be couched in terms whose application we can recognize. (99)

So, an option available to an agent in a situation—a hidden door behind the curtain, say—is no option at all if the agent centered at this situation does not recognize it as such. Gibbard does not to my knowledge define "subjective equivalence," but on one not unnatural take, one's doxastic centered possibilities are subjectively equivalent—these centered worlds are, for all the agent can tell, who they are. But if so, then a hyperplan cannot issue distinct verdicts on any pair of centered worlds left open by given a state of belief, because these will present the same options. This would remove some of the motivation for modeling

24 Of course, if the options are fixed by what is doxastically possible for a centered agent, then if an agent has exactly the same doxastic possibilities open at each of her doxastic possibilities, then for all centered worlds  $c, c'$  in any prosaically factual belief state  $B$ ,  $O_c = O_{c'}$ . So, the problem won't arise. But then it's not clear why we need to model in terms fact-plan possibilities at all, as contrasted with the alternative model I am about to suggest.

normative thinking and factual believing together in terms of a single set of fact-plan alternatives: after all, if a hyperplan cannot issue distinct verdicts on any pair of centered worlds left open by given a state of (factual) belief, then it doesn't matter which hyperplans a plan-laden belief state pairs with which centered worlds. You might as well have two separate sets, a set of centered worlds (factual belief) and a set of hyperplans.

Moreover, although a hyperplan is a function on centered worlds, it appears a hyperplan is only sensitive to one feature of a centered world: the subjective predicament of the agent at the center. That suggests we might just as well view hyperplans as functions on subjective predicaments. If we think of a subjective predicament as representable by a body of information—a set of centered worlds—then a natural idea would be to reconstrue hyperplans as functions on sets of centered worlds, rather than on centered worlds.

Let me assemble these considerations into a concrete proposal. Suppose now that hyperplans are functions on states of information rather than centered worlds: they are *information-based*. Associate Holmes with two sets:  $B$  (a set of centered worlds—his *doxastic state*, a kind of state of information) and  $H$  (a set of hyperplans capturing his views about what is permissible relative to various predicaments). Assume that any state of information, like Holmes's belief state  $B$ , fixes a set of options  $O_B$ . Take it the hyperplans in  $H$  speak to what to do given the options fixed by the whole of  $B$ . No longer do we have any special pairing of particular hyperplans with particular doxastic alternatives. Then the idea is to use these components to give Plan B+'s truth-conditions for (3), as follows:

For all  $h$  in  $H$ :  $h(B)$  permits only the options in  $O_B$  that entail packing by the agent the option is centered on.

Informally, this says that Holmes thinks that he ought to pack when all of the hyperplans left open by his normative state require packing when evaluated relative to his state of factual information. Observe that on this proposal, the truth of (3) owes both to  $B$  and  $H$ ; again, we have a kind "hybrid" view.

Let me offer two reasons to be interested in this way of modeling *ought* thoughts.

### 3.2. Nonpersistence of Ought Thoughts

First, this proposal respects the apparent nonpersistence of *ought* thoughts: the fact that *ought* judgments seem capable of coming and going under the impact of strict information gain. Suppose the following are the case at  $t$ :

- (6) Holmes thinks he might still be able to make the train.
- (7) Holmes thinks that if he might be able to make the train, he ought to pack.
- (8) Holmes thinks that if it's too late to catch the train, it's not the case that he ought to pack.

So of course, at  $t$ ,

- (3) Holmes thinks that he ought to pack.

Now at  $t + 1$  Holmes learns it's too late to catch the train. So, at  $t + 1$ :

- (9) Holmes thinks that it's not the case that he ought to pack.<sup>25</sup>

It seems:

- From the point of view of his purely factual belief state, Holmes underwent a strict information gain from  $t$  to  $t + 1$ .
- From the point of view of Holmes's pure normative state—his views about what is allowed in various possible situations—Holmes didn't change from  $t$  to  $t + 1$ .

Of course, Holmes goes from thinking he ought to do something to thinking he needn't. Putting it that way makes it look like he underwent a normative change of mind. But I think we can make out a level of description where we can say that he didn't really undergo any change of normative opinion. What changed is his view about the world, and hence which aspect of his (stable, unchanging) normative view speaks to the situation he takes himself to be in.

The truth-conditions for (3) supplied by Plan A do not get this right. If Holmes's state at  $t$  is such that each fact-plan possibility it leaves open calls for packing, then no shrinking of this set moves Holmes to a new state where packing isn't the thing to do. Some kind of plan-laden belief revision happened; the state was replaced with an entirely new set of fact-plan possibilities. But this seems an unnatural way to model a case of strict gain in information.<sup>26</sup> Plan B+, by contrast, seems to fit the facts. (The idea that *ought* thoughts might come and go under the impact of strict information gain is a main theme of Kolodny and MacFarlane (2010), and it is built into their semantics for deontic modals.)

One way to put pressure on Plan A here is to ask: what supposed condition on fact-plan alternatives obtains throughout Holmes's state if (8) is true? It is not easy to discern a natural answer. Do we check what is to be done relative to fact-plan alternatives in Holmes's state where it's too late to catch the train (the alternatives rendering the antecedent true)? But if we are assuming Holmes thinks he ought to pack and assuming Plan A's conception of what that means, then we already know such antecedent alternatives require packing, and hence are in tension with the consequent of (8).

<sup>25</sup> Hold fixed the time the packing is supposed to be happening, according to Holmes, across (3) and (9)—the same thing thought is at issue.

<sup>26</sup> One could try appealing to the changes in Holmes higher-order beliefs that take place between  $t$  and  $t + 1$ . How best to think about that will interact with the question how *de se* updating gets handled.

Plan B+, in contrast, would allow one to say the following. If (8) true, it is because when you start with Holmes state of information  $B$  and strictly add to it the information that it's too late ( $B \cap \text{TOO LATE}$ ), all of Holmes's hyperplans are such to prohibit packing relative to the options this updated state of information fixes.<sup>27</sup>

An important thing to notice here is the following: if you take Holmes's doxastic state and intersect it with the set of centered worlds where it's too late to catch the train, the resulting set is not actually a possible belief state.<sup>28</sup> Yet it does not seem incoherent to suppose that this body of information nevertheless fixes a set of options. Reflection on conditional thinking seems to suggest we need the idea that a set of options can be fixed by a state of information that is not itself a possible doxastic state. This seems to encourage to move beyond the Plan A idea that the options that hyperplans interface with are always fixed by some possible subjective state of an agent.

### 3.3. Other-Locating Deontic Thinking

A second consideration in favor of Plan B+'s conception of the truth-conditions of things like (3) comes in when we think about how to model one's views about how others are permitted or required to act. We could separate two kinds of such thinking:

- **Self-locating deontic thinking.** Thinking what is to be done as if in the subjective circumstances of another (taking on their beliefs, desires, and all).
- **Other-locating deontic thinking.** Thinking what is to be done when in the position of another, but in the world as one (not the other) takes it to be.

Other-locating deontic thinking seems to me hard to model along the lines of Plan A. But there is something natural we can say on the information-sensitive approach. A case will help us fix ideas (discussed in Yalcin 2012):

John's puppy has been poisoned; so too Niko's kitten. There is only enough antidote left to save one of their pets, but the price is too high. Jay know all this. His view is that the thing for John to do, given his situation, is to steal the last of the antidote and give it to his puppy. Jay also thinks that the thing for Niko to do, given his situation, is to steal the last of the antidote and give it to his kitten. That is:

- (10) Jay thinks Niko ought to steal the remaining antidote and give it to his kitten.
- (11) Jay thinks John ought to steal the remaining antidote and give it to his puppy.

<sup>27</sup> Of course, one eventually requires a theory of conditionals that jives with this conception of the truth-conditions of (8). See, for example, Yalcin (2007), Kolodny and MacFarlane (2010), and Gillies (2010).

<sup>28</sup> Assuming any relation of doxastic accessibility obeys the axioms K and J (where J:  $\Box\phi \rightarrow \Diamond\Box\phi$ ). The trouble is that  $B \cap \text{TOO LATE}$  of course entails  $\text{TOO LATE}$ , but the centered worlds in  $B \cap \text{TOO LATE}$  are centered on an agent that does not believe  $\text{TOO LATE}$ . Thus an agent in the putative belief state  $B \cap \text{TOO LATE}$  would (1) believe  $\text{TOO LATE}$  and yet (2) believe she doesn't believe  $\text{TOO LATE}$ .



Take it Niko and John can't *both* steal the antidote, and Jay knows that. But Jay's view seems coherent. How to capture this?<sup>29</sup>

Suppose  $H$  is Jay's normative state and  $B$  is Jay's ordinary factual-but-self-locating doxastic state. Let  $B_N$  be the state that comes from  $B$  by shifting all the centers to Niko.<sup>30</sup> Likewise,  $B_J$  comes from  $B$  by shifting the centers to John. Then we can say the following:

- (10) is true iff for all  $h$  in  $H$ :  $h(B_N)$  permits only options in  $O_{B_N}$  that entail antidote-stealing for Niko's kitten.
- (11) is true iff for all  $h$  in  $H$ :  $h(B_J)$  permits only options in  $O_{B_N}$  that entail antidote-stealing for John's puppy.

These can both be true. The point here is that Plan B+ shows us how Jay's state of mind is coherent.

A notable feature of the story, which we saw already with conditionals at the end of the last section, is that  $B_N$  and  $B_J$  are not even possible belief states (see fn. 14). Still, it seems natural to consider these "supposable" bodies of information as each fixing a set of options. We seem to have an ability to think what to do relative to bodies of information that are only hypothetically entertainable—bodies of information that could never even possibly correspond to a belief state.

This is a good time to ask: is there not also a Plan A+? Couldn't we package this new-fangled conception of hyperplans with Gibbard's original style of metatheory? I am not so sure, and the reason is to do with the notable feature just referenced. Gibbard's metatheory wants to explain hyperplans as the states of maximally decided planners—hyperplanners. This aspect of his metatheory is part of what drives him to model with fact-plan possibilities. But it is hard to see what would correspond in his metatheory to the information states needed by Plan B+, in particular the ones that don't correspond to any possible doxastic state. One of the lessons of the cases that motivate Plan B+ is that not all thinking what to do can be theorized in terms of hyperdecided states being ruled in or out.

#### 4. Closing

I hope those who rejected Plan A because of its metatheoretic picture, or because of its seeming ambition to reduce normative thought to planning, are interested to see a related view, Plan B, without these features. Plan B fits into a broadly 'representationalist' conception

<sup>29</sup> On the face of it, this would seem to be a challenge for a textbook deontic logic, one appealing to the idea of *what is true in the worlds viewed as ideal* according to the agent. In the worlds ideal according to Jay, is it Niko or John who steals the antidote?

<sup>30</sup> Or Jay could have an individual concept or role associated with Niko, and we could use that to find the person Jay takes to be Niko at his doxastic possibilities. I skip over this complication.

of the mental, though expressivism is sometimes viewed as a competitor to that position. I hope also that Plan B+ draws out the point that among those who enjoy modeling with hyperplans, there remains nontrivial space for intramural debate about the details—and that debate can help to inform metatheoretic questions.

One thing I haven't discussed is the Frege-Geach problem. I have focused on models of normative judgment and the ways these can be philosophically glossed, and not on the compositional semantics or pragmatics of normative language. But I often put Plan A and Plan B as views about the truth-conditions of attitude ascriptions embedding normative vocabulary. In this way, we were implicitly generating constraints—racking up debt, you could say—in semantics-pragmatics. As I see it, both Plan A and Plan B need eventually to vindicate their proposed truth-conditions compositionally. (Or show that compositionality doesn't matter—as expressivists who take a Horwichian path might.) This is one of the core aspects of the Frege-Geach challenge. Which view has an easier time here? Is either approach even remotely plausible on linguistic grounds? These are natural next questions.<sup>31,32</sup>

## References

- Ayer, Alfred J. (1936). *Language, Truth and Logic*, 2nd ed. New York: Dover.
- Blackburn, Simon (1993). *Essays in Quasi-Realism*. Oxford: Oxford University Press.
- Bratman, Michael (1987). *Intention, Plans, and Practical Reason*. Cambridge: MIT Press.
- Carnap, Rudolf (1935). *Philosophy and Logical Syntax*. London: Kegan Paul.
- Charlow, Nate (2015). "Prospects for an Expressivist Theory of Meaning." *Philosophers' Imprint* 15, no. 23: 1–43.
- Copp, David (2001). "Realist-Expressivism: A Neglected Option for Moral Realism." *Social Philosophy and Policy* 18, no. 2: 1–43.
- Dandelet, Sophia (2017). "Plans, Preference, and Normative Judgment." Unpublished draft.
- Gibbard, Allan (1981). "Two Recent Theories of Conditionals." In *Is: Conditionals, Belief, Decision, Chance, and Time*, edited by W. Harper, R. Stalnaker, and G. Pearce, 211–47. Dordrecht: D. Reidel, 1981.
- (1986). "An Expressivistic Theory of Normative Discourse." *Ethics* 96, no. 3: 472–85.
- (1990). *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.
- (2003). *Thinking How to Live*. Cambridge, MA: Harvard University Press.
- (2006). "Reply to Critics." *Philosophy and Phenomenological Research* 72, no. 3: 729–44.
- (2012). *Meaning and Normativity*. Oxford: Oxford University Press.

<sup>31</sup> Yalcin (2018) contains some discussion.

<sup>32</sup> Thanks to audiences at the University of Toronto, Rutgers, and the 2017 Philosophy Mountain Workshop for helpful feedback on earlier versions of this work. Thanks also to Sophie Dandelet, William Dunaway, Melissa Fusco, Anandi Hattiangadi, Joshua Petersen, and David Plunkett. I am especially indebted to conversations with Alejandro Perez Carballo for discussion of the modeling questions that arose in section 3.

- Gillies, Anthony. (2010). "Iffiness." *Semantics and Pragmatics* 3, no. 4: 1–42.
- Hare, Richard M. (1981). *Moral Thinking: Its Method, Levels, and Point*. Oxford: Clarendon Press.
- Heim, Irene (1992). "Presupposition Projection and the Semantics of Attitude Verbs." *Journal of Semantics* 9, no. 3: 183–221.
- Hintikka, Jaakko (1962). *Knowledge and Belief: An Introduction to the Logic of Two Notions*. Ithaca, NY: Cornell University Press.
- Horwich, Paul (1998). *Truth*, 2nd ed. Oxford: Clarendon Press.
- Kolodny, Niko, and John MacFarlane (March 2010). "Ifs and Oughts." *Journal of Philosophy* CVII, no. 3: 115–43.
- Lassiter, Daniel. (2011). *Measurement and Modality: The Scalar Basis of Modal Semantics*. PhD thesis, PhD dissertation, New York University.
- Levinson, D. (2003). "Probabilistic Model-Theoretic Semantics for 'want'." In *Proceedings of SALT XIII*, vol. 13, edited by R. Young and Y. Zhou, 222–39.
- Lewis, David K. (1979). "Attitudes De Dicto and De Se." *Philosophical Review* 88, no. 4: 513–43.
- (1988). "Relevant Implication." *Theoria* 54, no. 3: 161–74.
- (1994). "Reduction of Mind." In *A Companion to the Philosophy of Mind*, edited by S. Guttenplan, 412–31. Oxford: Blackwell.
- Moss, Sarah (2018). *Probabilistic Knowledge*. Oxford: Oxford University Press.
- Quine, Willard V. (1948). "On What There Is." *Review of Metaphysics* 2, no. 1: 21–38.
- Railton, Peter (1992). "Nonfactualism about Normative Discourse." *Philosophy and Phenomenological Research* 52: 961–68.
- Ridge, Michael (2014). *Impassioned Belief*. Oxford: Oxford University Press, 2014.
- Rosen, Gideon (1998). "Blackburn's *Essays in Quasi-Realism*." *Nous* 32, no. 3: 386–405.
- Russell, Bertrand (1935). *Religion and Science*. Oxford: Oxford University Press.
- Santorio, Paolo (2016). "Nonfactual Know-How and the Boundaries of Semantics." *Philosophical Review* 125, no. 1: 35–82.
- Scanlon, T. M. (May 2006). "Reasons and Decisions." *Philosophy and Phenomenological Research* LXXII, no. 3: 722–28.
- Schroeder, Mark (2008a). *Being For: Evaluating the Semantic Program of Expressivism: Evaluating the Semantic Program of Expressivism*. Oxford: Oxford University Press.
- (2008b). "What Is the Frege-Geach Problem?" *Philosophy Compass* 3, no. 4: 703–20.
- (2009). "Hybrid Expressivism: Virtues and Vices." *Ethics* 119, no. 2: 257–309.
- Stalnaker, Robert. (1976). "Propositions." In *Issues in the Philosophy of Language: Proceedings of the 1972 Oberlin Colloquium in Philosophy*, edited by Alfred F. MacKay and Daniel D. Merrill, 79–91. New Haven: Yale University Press.
- (1984). *Inquiry*. Cambridge, MA: MIT Press.
- (2014). *Context*. Oxford: Oxford University Press.
- Starr, William (2016). "Dynamic Expressivism about Deontic Modality." In *Deontic Modality*, edited by Nate Charlow and Matthew Chrisman, 355–94. Oxford: Oxford University Press.
- Willer, Malte (2017). "Advice for Noncognitivists." *Pacific Philosophical Quarterly* 98, no. S1:174–207.

- Yalcin, Seth (2007). "Epistemic Modals." *Mind* 116, no. 464: 983–1026.
- (2011). "Nonfactualism about Epistemic Modality." In *Epistemic Modality*, edited by Andy Egan and Brian Weatherson, 295–332. Oxford: Oxford University Press, Oxford.
- (2012). "Bayesian Expressivism." *Proceedings of the Aristotelian Society* CXII, no. 2: 123–60.
- (2018). "Expressivism by Force." In *New Work on Speech Acts*, edited by Daniel Fogal, Daniel Harris, and Matt Moss, 400–29. Oxford: Oxford University Press.