HOW TO BE (AND HOW NOT TO BE) A NORMATIVE REALIST:

A DISSERTATION OVERVIEW THAT ASSUMES AS LITTLE AS POSSIBLE ABOUT MY READER'S PHILOSOPHICAL BACKGROUND

Consider the question, "What am I going to have for breakfast?" One way to think about this question is as similar to the question, "What is Bob Hoskins going to have for breakfast?" Both questions are about what someone is going to have for breakfast. And in both cases there is presumably an answer, just as there is an answer to any question about what is going to happen in the future. But the first question is more naturally read not just as a question about what is going to happen, but as a *deliberative* question, a question about what you are going to *decide* to have for breakfast.

In this "deliberative mode," you might ask, further, whether one decision is in any sense superior to another, whether there are facts about what you *ought* to do, or about what you have *reason* to do. To an extent, this might be a moral question. You might believe that certain breakfasts would be *wrong* to have, perhaps (controversially) those that include meat or (less controversially) those that include your next-door neighbor. But even setting such obviously moral questions aside, you may have further deliberating to do. For example, there might be certain things that you *want* for breakfast, and you might think that this desiring is relevant to what you *should* have for breakfast. Or you might think that other things are important, such as that you remain thin, and that this is also relevant to making the *correct* choice.

Facts about what you should, ought, or have reason to do, along with (apparently relevant) facts about what is good, important, or valuable are what we call *normative* facts. Most people think that there are, quite obviously, lots of normative facts: moral facts about what is right and wrong, prudential facts about what is good or bad *for you,* knowledge-relevant facts about what you should or should not believe, and so on and so forth.

But though there is wide agreement that there are such facts, (at least) two noteworthy things raise interesting philosophical problems. First, there is widespread *dis*agreement about what the particular facts *are*. And, second, normative facts seem to be quite unlike other kinds of facts. Consider: Presumably, the world is a particular way. Non-normative (or "descriptive") facts are, on one natural view, just propositions that accurately describe what the world is like. The world is a particular way, and that explains why 'the cat is on the mat', '2 + 2 = 4', or 'today is Sunday' are all true (if they are). But normative facts (apparently) don't describe, rather they *prescribe*; they are facts about how things should be, not how they are. And so it seems that even if we had a full description of how the world is, we would need more to explain why claims like 'lying is wrong', 'you should believe only true things', or 'the unexamined life is not worth living' are true (if they are).

Not surprisingly, these philosophical issues are the purview of philosophical ethics or, more broadly, normative theory. More specifically, the problem of how there could be normative facts is a problem for *metanormative* theory, which is (roughly) the branch of philosophy that deals with second-order questions about the *nature* of normative facts, knowledge and language (as opposed to normative theory, which is (roughly) concerned with what the normative facts actually are). My dissertation is in metanormative theory (with a bit of normative theory at the end). It is about precisely the question discussed above: What sort of facts would normative facts be?

More specifically, my dissertation is a critical examination of the most prominent contemporary answers to this question. Broadly and somewhat roughly speaking, we can divide modern metanormative theorists who think there are normative facts into three relevant camps: the expressivist, the non-naturalist and the naturalist. In each of three chapters, I discuss one of these views. I conclude that our best bet is a view most closely related to naturalism. In what follows, I'll say a bit about what that all means, and what kind of arguments I offer.

Expressivism

Normative facts have another interesting feature that makes them (apparently) different from descriptive facts—they seem to have a special link with motivation. Often (though not always) if you judge that you *should* do something, you are motivated to do it (this doesn't mean you *do* do it, just that you feel some "pull" in that direction). Part of what makes this interesting is that there is an entrenched philosophical view that it is our *desires,* and not our *beliefs,* that motivate us. You can tell me anything you want about how the world is: the cat is on the mat, 2 + 2 = 4, today is Sunday. Until I have a relevant desire—to go pet the cat, to evenly divide four pieces of pizza, to go to church on Sunday—it doesn't seem like these facts are going to motivate me to do anything. But again, normative judgements seem different. When we judge that going to church is the *right* thing to do or what we *ought* to do, we are thereby motivated (again, perhaps just a bit and perhaps not always) to actually *go* to church, even if we don't want to.

Combining this with a number of other things—notably, the difficulty of making sense of normative facts in the first place, plus the "feel" of normative judgement (moral judgements in particular often come with a kind of "emotional force" that judgements like 'the cat is on the mat' do not)—a number of philosophers have proposed that, surprisingly, normative judgements aren't beliefs after all, but are more like desires. On their view, when I say 'lying is wrong', I'm not expressing a belief about how the world is (or even should be). Rather, I'm expressing some desire-like, "non-cognitive" attitude. Historically, expressivists have talked about this using analogies. They say that 'lying is wrong' is akin to an expression like 'boo lying!', a command like 'don't lie!', or some emotional response to lying.

To most people, especially outside of philosophy, two things are immediately apparent about this view. The first is that it just doesn't sound right. "I really do have normative beliefs!" you might think, "I think that lying is wrong and by that I really do mean that lying has this particular quality,

just like when I say that apples are red I'm saying that apples have a particular quality." Second, and perhaps even more obviously, this does *not* seem like a view on which there could be normative *truth*. After all, if 'lying is wrong' just means 'boo lying!', then apparently it's not the sort of thing that could be true or false.

Now, this raises a whole host of problems for expressivism, some of which can be dealt with fairly easily and some of which cannot. But there is one problem that is arguably expressivism's greatest challenge. The problem stems from the fact that normative language behaves, logically, just like descriptive language. Suppose I argue:

1. If the cat is on the mat, the cat is sleeping.

2. The cat is on the mat.

3. Therefore, the cat is sleeping.

This is a valid argument. This means that *if* 1 and 2 are both true, then 3 is true, too. Here is another apparently valid argument:

1. If lying is wrong, then saying that it's Sunday when it's Monday is wrong.

2. Lying is wrong.

3. Therefore, saying that it's Sunday when it's Monday is wrong.

Again, *if* 1 and 2 are both true, it seems that 3 must be, too. But notice that the following is *not* a valid argument:

1. If boo lying, then boo saying it's Sunday when it's Monday.

2. Boo lying.

3. Therefore, boo saying it's Sunday when it's Monday.

This just doesn't look right. 'Booing' isn't the sort of thing that can fit structurally into sentences in this way. So 'lying is wrong' can't be *just* like 'boo lying', because the former has a logical structure that the latter lacks.

Here is a related problem. The sentences 'lying is wrong' and 'lying is not wrong' are obviously inconsistent. They contradict each other. If you think one and I think the other, we disagree. But 'boo' and 'yay' don't work quite like this. If I boo a sports team and you cheer for it, there is an extended sense in which we disagree, but we have not really contradicted one another.

Anyway, the upshot of all this is that the expressivist has to be able to tell us how it is that normative claims express desire-like attitudes instead of beliefs, but also function logically just like normal declarative sentences that *do* express beliefs. And there is a huge philosophical literature full of attempts to do just that, along with responses from non-expressivists claiming that they have not succeeded.

Now, let's set this problem aside for a moment and return to the question of how expressivists can say that there is normative *truth*. Surprisingly, many expressivists think that they *can* talk about normative truth (at least in an extended sense, the details of which I won't go into). The first chapter of my dissertation is an attempt to combine this truth problem with the logic problem above. I argue that the things expressivists have to say in order to solve the logic problem make it look as though, whenever you make any judgement at all, that judgement has precisely the sort of special connection with motivation that, above, I suggested that normative judgements—judgements that things are right, wrong, good, bad, etc.—have. On the assumption that *only* normative judgements have this kind of connection with motivation, it follows that all judgements are normative judgements. One problem with this, I argue, is that it makes it impossible for expressivists to understand what *nihilists* are saying when they judge things. Nihilists, by definition, don't make any normative judgements. And this means, that for expressivists, they can't make any judgements at all.

Now, this is only a problem if nihilism is a view that expressivists have to take seriously. What I argue is that expressivists have to take this view seriously *if* they want to maintain that they

are really making good on the idea that there is "objective" normative truth—that there are facts about what is good, bad, right, wrong, etc., rather than just facts about what normative attitudes we have. Consider an analogy: Right now, you are looking at a screen or a piece of paper. You are having a screen-like or paper-like experience. And you might think that *that you are having such an experience* is totally beyond doubt. Nevertheless, you can certainly wonder whether there is really a screen or some paper "out there." After all, maybe it's all a hallucination!

Anyway, the point is that you can entertain a kind of *nihilistic doubt* about whether the object of your apparent perception is really there. And that's because that object is seemingly external to you, an *objective* part of reality. My claim is that if normative facts are also supposed to be an objective part of reality (or even just might be), then we have to be able to doubt that they are there, and if we can doubt that they are there, then nihilism is at least a possible view. The upshot, then, is that expressivists can solve the logic problem only by denying objectivity. And that means they're not a real contender in the search for objective normative truth.

Non-Naturalism

Non-naturalism is perhaps the "default" metanormative view, in that it essentially just takes everything I said in my opening paragraphs about the distinctive nature of the normative and runs with it. Non-naturalists assert that there are objective normative facts, and that these facts really are totally independent of and unlike any descriptive facts. One way of picturing this is that there is a sort of non-physical "realm" of normative facts. There is a totally autonomous and unique Property of Goodness, and it "attaches itself" to certain bits of the descriptive world and thus makes them good in the sense that means they are relevant to what we ought to do.

Now, this way of describing the view might make it sound a bit silly (and not every non-naturalist would be okay with you picturing his view this way). But it is nevertheless an attractive

view precisely because it makes good on the comments I made in the opening paragraphs. After all, if we really want objective normative truth, what's the alternative? Normative facts seem totally unlike descriptive facts. They aren't just "made true" by the way that the world is. So if there really are normative *facts* (not just whatever "truth" the expressivist can squeeze out of his view) then it seems we must take this literally and suggest that the way the world is descriptively indeed does not determine what's good, bad, etc., but that we need actual Goodness out there to do that.

The problems for this view are fairly straightforward. If this stuff is abstract and non-physical, then how does it link up with descriptive stuff to make it good or bad? And how did we come to know and talk about it? It's not like we could have come into contact with the Property of Goodness the way we come into contact with tables and chairs. So non-naturalists face heavy explanatory burdens. They need to explain the connection between the normative and the non-normative, and in particular the connections between normative beliefs and normative facts that give us normative *knowledge,* and the connections between normative words and normative properties that let us *refer* to those properties.

Unfortunately, it turns out that on the non-naturalist view, normative stuff can't explain non-normative stuff, and non-normative stuff can't explain normative stuff. And once you realize that, it's hard to see how the non-naturalist is going to offer all those explanations I just said he had to offer. If neither the normative nor the non-normative can explain the other, how can we explain connections between them?

Indeed, it turns out that non-naturalists simply *can't* explain connections between the normative and the non-normative all the way down. At some point, they have to stop and say something like, "these things are good, those things are bad, and that's that." At this point, there is a debate to be had about whether this is possible at all. Some people think that if two things are distinct—if they really are two separate things (as the non-naturalist says the normative and non-

normative are)—there must be some explanation for any robust connection between them. Kindness (say) can't *just be* good if goodness and kindness are two different things; something has to explain *why* kindness is good. But not everyone agrees that an explanation is required, and this is not a debate I want to try to decide on, so let's just assume that non-naturalists are allowed to have some inexplicable normative facts.

Regardless, even non-naturalists admit that *some* connections are just not the sort of thing that could be unexplainable. For example, no one is comfortable saying that our normative beliefs just *happen* to by-and-large be true. So non-naturalists still owe us something. Unfortunately, their position here is apparently made worse by the fact that we already *know* something about what explains why we have the normative beliefs we do—it seems to be a combination of upbringing, community, evolutionary forces, and the like. If *those* things explain why we have the normative beliefs we do, then why think that they are capturing some truth about this mysterious Property of Goodness, rather than just being reflections of normative practices or something like that? If that's right, you'd better either give up on normative truth or accept that normative truth isn't objective, but is, say, somehow constructed out of our practices.

Here, it turns out, the non-naturalist has a sort of sneaky move he can make. Suppose, for instance, that one of the inexplicable connections between the normative and the non-normative is that community is good. Now suppose it also turns out that our normative beliefs come from our community, and that they serve the community. Well now, it might turn out that our normative beliefs are true! Why? Well, they promote community. Community is good. So they promote what's good. And it might well be that the best way for us to promote what's good is for us to have true beliefs about what's good. In short, it turns out that if we start by assuming certain connections between the normative and the non-normative, those might create a sort of "pre-established

harmony" between what's actually good and what we think is good. (This is a very rough sketch of a more complicated story some philosophers tell.)

Anyway, that's the sort of move non-naturalists have to make, both in the knowledge case and the one about how our words refer to actual goodness, etc. As you can probably tell, the story is going to be convoluted and maybe not all that plausible. Nevertheless, there might be important reasons to accept it, especially if it's the only way to make sense of the possibility of normative truth.

In the dissertation, I say a bit more about why it is that I think this non-naturalist story isn't very satisfying. But for the purposes of the larger project, all we really need is this: The non-naturalist story is convoluted and implausible enough that it looks like non-naturalism should probably be a last resort. So if there are any *other* prospects for making sense of normative truth, we should probably examine them thoroughly before going with non-naturalism. And this brings us to our final contender, naturalism.

Naturalism

Unlike expressivists, naturalists think that normative judgements really are beliefs, and that they concern (possible) normative facts. But unlike non-naturalists, naturalists think that the impression we have that normative facts have to be totally unlike descriptive ones is an illusion. It turns out, on the naturalist view, that properties like goodness really aren't all that different from normal descriptive ones like redness or sleepiness. On one way of thinking about it, given all the facts about how the world is, we have enough to figure out how it should be. Unfortunately, aside from the fact that this view can just "feel" wrong (in one common way of putting things, the gap between "is" and "ought" seems impossible to bridge), naturalism faces some serious challenges.

Time for a bit of history. Although there have been many philosophers who could be labeled "non-naturalists," the first philosopher to distinguish the view in these terms was a man named G.E.

Moore, around the beginning of the 20th century. Moore thought that non-naturalism had to be the right view. And he thought that he had a pretty good argument that naturalism was false.

To understand the argument, we need to understand the distinction between an "open" and a "closed" question. Think about the concept 'bachelor'. In traditional philosophy of language, concepts like 'bachelor' can be *analyzed*—we can give necessary and sufficient conditions for being a bachelor. Arguably, something is a bachelor if and only if it is an unmarried man (or something quite close to this).

One way to test whether a proposed analysis is correct or not is with an "open question" test. Suppose I tell you that Tom is a bachelor. I then ask you whether he is unmarried. Notice that the answer is obviously "yes." But it's not merely that the answer is yes, but that it seems that the answer is settled once I tell you Tom is a bachelor. Indeed, if someone were to reply "no," you would probably think that they had misheard you, or that they didn't really understand the question, or understand what "bachelor" means. Thus, this question seems "closed." And this indicates that the connection between 'bachelor' and 'unmarried' is a *conceptual* one.

Now contrast this with a case where, after telling you that Tom is a bachelor, I ask you whether Tom is friendly. You might know Tom well and know that the answer to this question is also "yes." But, importantly, whatever the answer is, it's not settled just by Tom's being a bachelor. If someone answered "no" to *this* question, you might think that person was *wrong,* but you wouldn't assume they were in some way confused about the question. So, this question seems "open." This indicates that the connection (if there is any) between 'bachelor' and 'friendly' is *not* a conceptual one.

Now here comes Moore. He suggests that we consider various similar questions using normative concepts like 'good'. Suppose you think that kindness is good, such that anything that's kind is therefore good. Then you ask: "I know that X is kind, but is X good?" Moore thinks (and

almost everyone agrees) that this question is open. Again, that doesn't mean that the answer isn't "yes" (indeed, I just asked you to suppose that you think it is); it just means that saying "yes" is more like affirming that Tom the bachelor is friendly than that Tom the bachelor is unmarried. And, says Moore, no matter what normative/non-normative concept pair you pick, the question will seem open.

This is the "open question" test, and it is supposed to indicate that there are no connections between normative and non-normative concepts, and that no normative concepts have non-normative analyses. Importantly, this *further* means (on the view Moore and everyone else at the time held) that normative terms don't *refer* to anything non-normative. So the upshot is supposed to be that normative properties can't just turn out to be non-normative, "descriptive" ones. And that's supposed to prove that naturalism is false.

Since Moore, a lot has happened in both the philosophy of language and metaphysics (not to mention metanormative theory). One thing that has happened is that the view just discussed, according to which what a term refers to depends solely on the correct analysis of the relevant concept, has been largely abandoned. One upshot of this is that we now know it is possible for two terms to refer to the same property even though there is no conceptual connection between them. For example, we know that our word "water" refers to H_2O, despite there being no conceptual connection between water and H_2O. (Notice that "This is a glass of water, but is it a glass of H_2O?" is an open question. If you're dubious, just imagine what someone in the 16th century would say about it.)

Without going into too many details, some philosophers have suggested that while Moore's argument rules out one kind of naturalism—the kind where normative and non-normative terms *mean* the same thing—it has not ruled out another—the kind where normative and non-normative terms *refer to* the same thing. Indeed, it has even been suggested that "good" might turn out to refer

to certain descriptive bits of the world in precisely the way that "water" has turned out to refer to H_2O.

On the canonical view, "water" refers to H_2O because (as we've discovered) H_2O is the stuff in the world that we've actually been talking about when we talk about water. In the terms traditionally used, it has turned out that H_2O is what *causally regulates* use of the term "water." Following this example, some (originally, a man named Richard Boyd) have suggested (more or less) that that "good" refers to whatever causally regulates our use of the term "good." Unfortunately for this view, it turns out that there are certain key judgements that we make about the "water" case that we don't make about the "good" case. Specifically, we tend to think that once we figure out what regulates our use of the term "water," we know what "water" refers to. But we do *not* think the same way about normative words like "good." There is a very famous and interesting thought experiment called "Twin Earth" that is meant to show this, which I have placed in an appendix for those who are interested in getting into the details a bit more here. If you are, skip down to the appendix. If you're not, just keep going on the assumption that there is a problem for naturalist views along these lines.

A number of philosophers think that this problem can be dealt with. And in the last section of the dissertation, I argue that there is hope. First, I suggest that there are certain claims about the nature of normative properties and normative truth that act as *constraints* on what those terms refer to. I then argue that the only way to construct a viable naturalist theory is to show that certain substantive facts about what we ought to do follow directly from a full understanding of the implications of these constraints. Explaining how this works exactly gets into some rather technical stuff, so I will leave trudging through it to those of you who want to read the actual dissertation. Here, then, let me close by just saying a few words about how I leave things.

I mentioned above that I dip into normative (as opposed to metanormative) theory in the dissertation. This happens in the final section of the last chapter, where I say a bit about how we might try to develop a naturalist view along the lines I've suggested. Here I'll say something even *more* sketchy about my sketch there.

There is a view in metanormative philosophy called "constitutivism." Constitutivists think that the normative facts are what they call "the demands of agency." We are agents—deliberating creatures—and (supposedly) there are certain rules that govern our deliberation and our action just in virtue of our *being* agents. Those rules (along with other non-normative information about the world), say the constitutivists, are all that we need to figure out what we ought to do.

Unfortunately, constitutivists face a kind of "open question" challenge of their own. It seems we need to know why we should be agents at all. Why not just not deliberate at all, or why not deliberate in opposition to the rules just mentioned? In other words, "I know this is a demand of agency, but do I have reason to do it?" seems like an open question. So in the final part of dissertation, I try to make some headway towards showing that given certain constraints on normative terms, it turns out that we actually *do* have reason to be agents. And if that's true, and the constitutivsts can make good on their promise of growing a full-blown normative theory out of that one seed, then we can have a full-fledged, naturalist view of normative truth. If not, then it seems we either have to keep plugging away at that non-naturalist story, or we give up on objective normative truth altogether.

Appendix: Twin Earth

Imagine that there is a world called Twin Earth that is exactly like Earth in every way but one: On Twin Earth, the stuff in the lakes, streams, oceans, etc. behaves (for all anyone can tell) exactly like water here on Earth, but it turns out that while the stuff here is H_2O, the stuff there is some other

compound, call it XYZ. The point of the original Twin Earth case was to show that when we think about our word "water" and the Twin-English word "water," people tend to think that they mean different things. If I got into an argument with my doppelganger on Twin Earth about whether I was drinking a glass of water, people tend to think that he and I are having a merely verbal disagreement—we're using similar sounding words to talk about different things and once we realize that *my* word "water" refers to H_2O and *his* refers to XYZ, the disagreement will be resolved.

Remember that we have stipulated that in every other way, the English and Twin-English words "water" are identical. Yet it seems their meanings differ. That tells us two things. First, it tells us that for some words, knowing what's going on in our heads and how we think about a term isn't enough to tell us what it refers to (this is what makes trouble for the view that I said Moore and others were taking for granted). Second, it tells us that in the case of some terms, like "water," facts about what causally regulates our use of that term seem to make a contribution to its meaning by determining (in part, at least) what it refers to. (Thinking about this in terms of the open/closed distinction, you might notice that "I know this is what causally regulates our use of the term "water," but is this water?" seems like a closed question—of *course* it's water (actually, there are complications here but hopefully the intuitive point is clear enough).)

Ok, so much for "water." What about "good?" Let's imagine a new Twin Earth. *Normative* Twin Earth is just like Earth in every way but one: Our use of the term "good" is regulated by things that make people happy (that doesn't seem true but assume it for the sake of argument) while their use of the term "good" is regulated by things that give people as much freedom as possible. According to the view under discussion, if "good" is like "water," then if I were to argue with my Normative Twin Earth doppelganger about whether something is good, we would have a merely verbal disagreement. That is, we should think that once we understand that *my* word "good" refers to happiness-promoting things and *his* refers to freedom-promoting things, we will recognize that we

were just using two words that sound similar but that we are really talking about different things and the disagreement will be resolved.

Most people don't think this, though. They think that if I were to argue with my doppelganger about whether something is good, we would be having a genuine normative disagreement. And finding out that our words are regulated differently would make no difference. We would just conclude that one (or both) of us has been referring to the wrong thing—something that's not actually good (at least not always). (Thinking about this in terms of the open/closed distinction, you might notice that "I know this is what causally regulates our use of the term 'good', but is this good?" seems like an *open* question.)

The point of this is that it looks like the theory that says that terms like "water" refer to whatever causally regulates their use doesn't apply to normative terms. And that means that at least that theory for how normative terms might come to refer to something descriptive is likely false. And just like in Moore's case, those who offer this argument think that they can construct a similar Twin Earth case to undermine *any* theory you might propose to vindicate naturalism.