Norm Descriptivism:
An Account of Normative Guidance and Inquiry

Howard Nye
June Ku

Perhaps the most basic challenge we face when we try to understand judgments about normative reasons for action is that of explaining how they play two apparently conflicting roles: (1) unlike most descriptive beliefs, they are intimately associated with motivational states that lead us to action, but (2) like descriptive beliefs, we inquire into their truth and falsity when we deliberate. In this paper we explore what more exactly these two roles amount to, including both the kind of connection normative judgments have to motivation and the kind of inquiry we undertake when we deliberate about what to do. We present a view of judgments about an agent's reasons for action according to which they are descriptive beliefs about deep features of that agent's psychology – namely the prescriptions of the most fundamental principles that she accepts. We argue that this view offers us the best explanation of both the connection normative judgments have to motivation and what goes on in deliberative inquiry.

1. Introduction

Our topic here will be judgments about an agent's *normative* or *justifying* reasons for action, or judgments about whether there are considerations that count in favor of the agent's doing something. Our topic should thus be distinguished from judgments about an agent's *motivating* or *explanatory* reasons, or judgments about which considerations answer the question "why did she do it?" whether or not they contribute to justifying her doing it. We should also clarify that our talk of judgments about reasons is not meant to focus on events of belief formation; it is simply intended as a way of referring to the mental states expressed by claims about reasons that is neutral between those who think they are descriptive beliefs and those who think they are something else.

Judgments about one's normative reasons for action, or what one has reason to do, appear to play two conflicting roles. On the one hand they seem – unlike descriptive beliefs – to be intimately associated with motivational states that lead us to action. It seems close to a platitude that judging that one has reason to do something tends to come along with motivation to do it, but merely contingent that any descriptive belief should be associated with a motivation. If, for instance, one happens to have a standing desire to feed hungry goats, then coming to believe that Bill the goat is hungry can cause one to be motivated to feed him. But coming to judge that one has reason to feed hungry goats seems capable of motivating one to feed them in the absence of anything like a standing desire to do whatever one has reason to do.

On the other hand, judgments about what one has reason to do seem – just like descriptive beliefs – to be the kind of mental state that we can discover to be true or false through a kind of inquiry. As with our descriptive beliefs, we can wonder whether our judgments about

what to do are true, and we have procedures for determining whether or not they are. In general, when we inquire into whether or not *P* is true, we attempt to arrive at knowledge that *P* or that not *P*. But to know that *P* seems to require that *P*'s truth figures into the explanation of one's belief that *P*.[1] As such, normative inquiry seems to aim at a state where the truth of one's normative judgments plays a role in explaining why one holds them.

These apparently conflicting roles animate different metanormative accounts of the meaning of claims and the content of judgments about reasons for action. Non-cognitivist or expressivist accounts, which hold that these judgments are – and these claims express – non-cognitive attitudes, seem to do very well at explaining the first, motivation-oriented role. If the mental state expressed by the claim 'I have reason to *φ*' *just is* something like a motivation to *φ*, then it should be no surprise that we find ourselves with motivation to do something when we judge that we have reason to do it. But expressivist accounts seem to have a much more difficult time explaining what it is to wonder whether one has reason to do something and what one is up to when one goes about trying to figure out whether one does. Normative inquiry seems to aim at a state of normative knowledge in which the truth of one's judgments about what to do explains why one holds them, but it is difficult to see how the truth of a non-cognitive attitude (if this is even a coherent notion) could explain why one holds it.[2]

On the other hand, cognitivist or descriptivist accounts, which hold that judgments about reasons for action are – and claims about such reasons express – a specific kind of descriptive belief, seem to have the opposite virtue and vice. If judgments about what to do are representations of normative facts that can, like other descriptive facts, explain such phenomena as why we hold the beliefs we do, we can explain deliberation straightforwardly as aiming at a state in which the truth of our judgments about what to do explains why we hold them. But descriptivist accounts seem to have a much more difficult time explaining the intimate connection between our judging that we have reason to do something and our being motivated to do it. If judgments about what to do are just another kind of descriptive belief about the world - like the belief that snow is white or that there are more than five chairs in the room - it seems difficult to see how these judgments can be any more essentially connected to motivation than other descriptive beliefs.

---

[1] We mean to construe "figuring into an explanation" so as to include both those facts that are indispensable to the explanation and those facts that are analytically entailed by those that are indispensable. It seems plausible to suppose that the explanation of a fact must entail that it (or – in indeterministic cases – the probability that it) obtains, and that when someone knows something about the future there must be a common explanation of both her belief and what she knows (see for instance (Hempel 1965), (Railton 1978), and (Dretske 1981)). If this is right then our kind of requirement on knowledge can, like Dretske's (1981) account of knowledge as information caused belief, subsume what seems right about the causal theory of knowledge but also capture the explanatory relationship that must obtain between beliefs about the future and their truth for them to constitute knowledge. Moreover, because a fact's causing a belief is presumably not the only way the fact can enter into the ontic explanation of - or reason why it is the case that - one holds the belief, we think that this requirement can also cover the kind of explanatory relationship that must obtain between beliefs about necessary propositions and their truth for them to constitute knowledge. See also (Gibbard 2003), chapter 13 on the notion of "deep vindication" and knowledge in the "more demanding sense" for a related formulation.

[2] A role in the explanation of prosaically descriptive phenomena such as our having the attitudes we do is exactly the dividing line drawn by (Gibbard 2003, 183-185) between descriptive truths and facts and the normative truths and facts that an expressivist quasi-realist can claim to exist.

In this paper we will present a descriptivist metanormative theory of judgments about reasons for action that we think can overcome descriptivism's typical vice and explain the connection between judging that one has reason to do something and being motivated to do it. To get the intuitive idea behind our view, consider the following. To judge that an entity has reason to do something is different from simply judging that its doing it would be good or something we should hope for or promote. When volcanoes fail to erupt and kill people, they do something that we should hope for and (to the extent we can) promote, but it would be absurd to think that they do something they had reason to do. Similarly, when in fair competition with another agent she does something that gives her an advantage, she does something that one may have reason to hope she doesn't do and try to prevent her from doing, but nonetheless does something that she has reason to do.

A natural attempt to explain what is distinctive about judgments that an agent has reason to do something is to identify them with judgments that her doing it will promote her actual ends. The problem with this is that agents can deliberate about which ends to pursue, and take themselves to have reason to do only what will promote the ends they *should* pursue.[3] But if this is right, the above kinds of considerations suggest that judging that an end is rational for an agent to pursue is quite a distinctive state – different, for instance, from merely judging that it would be good that she pursue it or that we should encourage her to pursue it.

What might seem distinctive about judging that an agent has reason to pursue an end is that it amounts to thinking that the agent could correctly reason her way to endorsing or pursuing the end, where correct reasoning is a process of going from what one accepts to what is genuinely prescribed by what one accepts. If this is correct, then to judge that an agent should pursue an end is to judge that the agent is in a sense already committed to pursuing it; that she already accepts principles that prescribe her pursuing it. Agents may of course accept conflicting principles, and it seems that they can reason their way to accepting some and rejecting others. But what goes for having reason to pursue ends and reasoning one's way to pursuing them goes for having reason to accept principles and reasoning one's way to accepting them. If an agent has reason to accept or reject a principle, then she must be able to go correctly from what she accepts to accepting or rejecting the principle. This, however, requires that she accepts something else even more deeply or fundamentally which prescribes that she accept or reject the principle in question.

What this leads to is a metanormative theory according to which to judge that an agent has reason to do something is to believe that her doing it is prescribed by the most fundamental norms or principles that she accepts. We call this view *Norm Descriptivism*. Our contention in this paper will be that Norm Descriptivism provides the best explanation of both the way in which deliberation about what to do is bound up with motivation and the kind of inquiry it is.

We begin in Section 2 by arguing that descriptivist views that cannot explain an essential connection between normative judgment and motivation are prone to deliver either an implausible form of error theory about reasons or an implausible analysis of normative concepts.

[3] This is at least a problem for taking the identification of judgments about what to do with judgments about what will promote one's actual ends to be *the whole* of the story of judgments about what to do. There may well be what we might call a "restricted" sense of 'reason to act' for which the foregoing identification is entirely correct. What we intend in the text is simply that there is also an unrestricted sense of 'reason to act' in which one only has reason to do what will satisfy those ends that one should pursue.

In Section 3 we consider a version of the Humean theory of practical reasons that we think can do better, but argue that it cannot plausibly explain how we revise our motives via reflective equilibrium methods when we inquire into what ends to pursue. In Section 4 we draw upon what the shortcomings of the Humean theory reveal about normative judgments to argue that deliberation is best explained as a process by which an agent attempts to figure out what is prescribed by the most fundamental norms she accepts. We extend our account to judgments about other agents' reasons in Section 5, where we contend, against versions of expressivism and relativism, that judgments about what others should do are judgments about what is prescribed by their fundamental norms, not just those one accepts oneself. We conclude by responding to three possible objections in Section 6, arguing that Norm Descriptivism can successfully explain our moral reasons for action and that it is in fact incoherent to think that an agent has reason to do something but that her doing it is not prescribed by the most fundamental norms she accepts.

2. A Dilemma for Judgment Externalism: Irrelevant Elimination or Non-Normativity

Above we suggested that a main problem with descriptivism and source of attraction to expressivism is that descriptivism seems driven to what we might call *judgment externalism*, or a denial of the *judgment internalist* thesis that it is a conceptual truth about judging that one has reason to φ that it tends to come along with motivation to φ.[4] Some descriptivists, however, might be inclined to think that this is not so much a problem for their views as a problem for the judgment internalist thesis. They might acknowledge that there is a connection between judgments about reasons and motivation, but deny that this connection is guaranteed by the content of these judgments or the kind of mental states they are. Just as it is a psychological fact about many actual mammals that appearances of snake-like features tend to make them fearful or averse, but no part of what it is to have such appearances to tend to have these responses, perhaps it is simply a psychological fact about actual agents that their judgments about what to do tend to motivate them.

We think, however, that such attempts to analytically detach normative judgments from motivation are unable to successfully locate our normative judgments among the many beliefs we hold. On the one hand, considerations of explanatory parsimony suggest that there are no descriptive but analytically irreducible normative facts, yet this seems to be a bad reason to embrace error theory about what to do. On the other hand, attempts by judgment externalists to analytically reduce normative facts to other kinds of facts seem open to objections reminiscent of Moore's "open question argument."

Explanatory parsimony gives us reason to think that there exist only those descriptive facts that either figure into our best explanation of the total phenomena or get analytically entailed by it. The former presumably include the facts discussed by fundamental physics, while the latter include facts about averages and (arguably) things like color, heat, chemistry, and psychology.[5] There must, however, be some constraints on what will for the sake of the parsimony principle be allowed to count as the "total phenomena", or it would have no teeth.

[4] This terminology is derived from (Darwall 1983, 54).

[5] The example of averages is Harman's (1977). For an excellent treatment of the case that these other kinds of facts are analytically entailed by our best explanation of what there is, see (Jackson 1998).

We might think, for instance, that if we can explain such phenomena as thunder and lightning, the misfortunes of dishonest traders and those who were inhospitable, and the emergence of the Balkan and Rhodope mountains without any reference to facts about Zeus, then we have reason to believe that there are no such facts. But what if the defender of Zeus-facts were to object that the *total* phenomena include such phenomena as Zeus turning into a bull and raping women, and we *do* need facts about Zeus to explain these?

A good response seems to be the following. Facts about Zeus are not only unnecessary for explaining such phenomena as thunder and lightning, the misfortunes of the dishonest and inhospitable, and the emergence of certain mountain ranges. They are also unnecessary for explaining why people believed that Zeus turns into a bull and rapes women, and why they believed all of the other things they believed about Zeus. The general lesson seems to be that considerations of parsimony dictate the following. If we do not need a certain kind of descriptive fact to explain anything else, and we can best explain all of our beliefs about such facts without invoking them (or an explanation that analytically entails them), then we should not believe that there are any such facts.[6]

If we need facts about reasons for action to explain anything, it seems that it must be something about agents' attitudes, behavior, or normative judgments. But to proximally explain agents' attitudes and behavior, we need at most their judgments about reasons; whether these judgments accurately correspond to facts about reasons is irrelevant. To explain agents' judgments about reasons, we need facts about their acculturation, which may be largely explained by the normative views of those around them. But follow the chain of acculturation back, and it looks as though you need explain only how tendencies to make certain normative judgments got passed down by mechanisms of biological or cultural evolution.[7] In the case of our capacities to form beliefs about such things as tables, chairs, and electrons, we need to posit the reliability of these mechanisms in tracking their subject matter to explain why they would enhance survival and reproduction and thus get passed down. But we do not need to posit the reliable tracking of irreducible facts about reasons by judgments about reasons - *in addition to* such judgments' directly tracking things like survival and reproduction - in order to explain the mechanisms by which we came to make judgments about what to do.

It looks, then, as though we will need to posit descriptive facts about reasons for action that were successfully "tracked" in evolution only if such facts are reducible to other facts that

[6] See for instance (Harman 1977) and (Gibbard 1990, 2003). David Enoch (2007) has recently objected to this criterion, arguing that it is enough if belief in a kind of fact is indispensable to a "non-optional" project for it to be the case that we should believe that it exists. Enoch says that by 'non-optional' he is unsure whether he means projects from which "we *can*not disengage, or rather those we should not disengage, or perhaps some combination of the two." We are quite unsure what Enoch means by a project 'we cannot disengage'; whether read as a claim about psychological, metaphysical, or conceptual impossibility it does not seem that either of his two examples – deliberation and explanation – really qualify. We are also quite unsure as to why it would be at all plausible to claim that simply because we should engage in project *P* and project *P* requires belief in facts of kind *F* that we have *epistemic* reason to believe in facts of kind *F*. But what really baffles us is how, given that he is a descriptivist and not an expressivist quasi-realist, Enoch can think that the deliberative project – that of figuring out what to do and why – is anything other than a sub-element of the explanatory project of figuring out what is the case and why.

[7] We doubt that anything really hangs on the details of the true story of how we came to make the normative judgments we do. Irreducible normative facts would seem just as superfluous had we been set up to make normative judgments by deities, or had we been spontaneously generated a few moments ago by lighting hitting a swamp, or whatever. We stick to the actual evolutionary story for heuristic purposes.

evolution did design our normative judgments to track. One might claim, for instance, that because descriptive facts about our reasons are identical to facts about what would maximize our pleasure, and evolution designed us to track facts about our pleasure, it designed us to track facts about our reasons. Such fact identities could be either analytic, like those between facts about brothers and male siblings, or synthetic, like those between facts about water and H_20.

The problem with holding that there are synthetic identities between descriptive facts about reasons for action and other kinds of facts is that such identities are explanatorily superfluous. As with other kinds of descriptive facts, we should only believe in facts about identities if they enter into (or are analytically entailed by) the best explanation of something - at the very least that of our believing in such identities. We should, for instance, believe that facts about water are identical to facts about H_20 because this enters into (or is entailed by[8]) our best explanation of what water is like and how we came to have the beliefs about it we do. But we should not, for instance, believe that facts about Mt. Olympus are identical to facts about Zeus's actual home because no such identity enters into (or get entailed by) our best explanation of how things are. An identity between facts about what to do and other facts like those about our pleasure would add nothing to our evolutionary story or any other part of our explanation of why people make the judgments about reasons for action they do. Explanatory parsimony thus entails that we should not believe that there are any such synthetic fact identities.

It seems, then, that we need neither irreducible nor synthetically reducible descriptive facts about reasons for action to explain our judgments about such reasons. As such, a judgment externalist who maintains that judgments about what to do are descriptive beliefs about such analytically irreducible facts is committed to saying that all of our beliefs about what to do are just as mistaken and untrue as beliefs about Zeus. But surely this is a terrible reason to embrace such error theory about reasons for action. The mere fact that we do not need a special kind of fact or fact identity to explain our beliefs about what to do does not seem to mean that nothing is really worth doing or that nothing can really count in favor of doing anything. So it looks like judgments about what to do cannot be beliefs about such analytically irreducible facts.

The only other option for the judgment externalist would be to hold that facts about reasons are *analytically* reducible to some other kind of facts that we do need for explanatory purposes. Some of the most plausible reductions might analyze judging that one has reason to do something as judging that one would be motivated to do it if one were under some particular, non-normatively specified conditions.[9] Or the judgment externalist might hold that to judge that one has reason to do something is to judge that one's doing it conforms to certain abstract rules (identified by their content) in the same way that this might be true of judging that one's actions conform to the rules of a game like Soccer.

The problem here is that tests reminiscent of Moore's "open question argument" suggest that these analyses are out of line with our intuitions about which judgments about reasons are coherent. It seems that for any descriptively specified conditions or set of abstract rules, one

[8] See (Lewis 1970) and especially (Jackson 1998) for a powerful case that this identity follows by analytic entailment from our best theory.

[9] For instance, a version reminiscent of Firth's (1952) account of moral judgments would hold that these conditions are those of full information and equal, vivid attention to all the (prosaically descriptive) facts, and a version reminiscent of Brandt's (1979) reforming analysis of judgments about rational desires would hold that these conditions are those of having undergone Brandtian "cognitive psychotherapy".

could coherently judge that one would be motivated to perform an act under the conditions, or judge that the act would conform to the rules, yet judge that one has no reason to perform the act. This is at least some evidence that judgments about reasons are not identical to these kinds of judgments, and puts an explanatory burden on those who would maintain such an identity.

We think that the best explanation of the vulnerability of any such judgment externalist theory to these problems is that, *simply in virtue of the kind of mental states they are*, judgments about one's reasons for action are intimately connected to motivation. If this is right, then descriptivist theories of judgments about reasons should seek to capture some form of judgment internalism. We now turn to a theory that we think draws much of its intuitive credibility from its apparent ability to do just this.

3. Analytic Humeanism and its Shortcomings

According to one reading of the Humean theory of practical reasons, which we shall call 'Analytic Humeanism', to judge that agent has reason to do something is to believe that her doing it will satisfy her non-instrumental motives, or bring about something she is motivated to bring about for its own sake.[10] In Section 1 we suggested that this might be an intuitively attractive way to explain what is distinctive about judging that someone has reason to do something, as opposed, for instance, to simply judging that her doing it would be good or something we should promote. We think that Analytic Humeanism's attractiveness in this regard may stem largely from its appearing well positioned to capture both of the apparently conflicting features of judgments about reasons that we mentioned at the outset of Section 1.[11] Whether an act will satisfy one's non-instrumental motives is a descriptive fact that one can hope to hook onto through inquiry. At the same time, simply in virtue of the kind of state it is, judging that one's doing something will satisfy a motivation one has tends to give rise to motivation to do it.

Unfortunately, Analytic Humeanism seems to commit us to an inadequate account of what goes on in deliberation about what to do. Such deliberation does not seem to be exhausted by attempts to determine what will satisfy our existing motives whatever they are. We seem to

[10] We here speak of 'motives' generally rather than 'desires' in particular because we think the most plausible versions of the Humean theory identify the acts an agent has reason to perform with those that will satisfy her non-instrumental motives whether or not these are desires in the ordinary English sense. For instance, if we have reason to bring about what we intrinsically desire then surely we also have reason to prevent what we are intrinsically averse to (where it seems rather implausible to think that aversion to *E* is identical to ordinary English desire that *E* not happen). Using the term 'desire' in a "thin" sense to simply express MOTIVATION would be harmless enough, but we wish to avoid any equivocation between this and the thicker, ordinary English sense.

[11] Another source of attraction to Analytic Humeanism might be that, as we suggested in note <<3>>, there may be a restricted sense of 'reason to act' of which the Analytic Humean thesis is true, which is easily confused with the unrestricted sense of 'reason to act' that we mean to be discussing in the text. We think, however, that Analytic Humeanism remains attractive as a thesis about all important normative senses of 'reason to act', even after one recognizes a distinction in principle between an unrestricted sense which depends upon the rationality of the non-instrumental motives an act serves and a restricted sense that does not. Analytic Humeans can either deny the coherence of unrestricted reasons judgments on the grounds that judgments about the rationality of non-instrumental motives are incoherent, or they can contend that the Analytic Humean thesis is true of both restricted and unrestricted judgments on the grounds that it's an analytic truth that the non-instrumental motives one should have are the non-instrumental motives one does have.

be able to evaluate our non-instrumental motivations *themselves* as rational or irrational, and take ourselves to have reason to do what will satisfy only those motivations that we deem rational.[12] Moreover, we seem to have a method of determining which non-instrumental motives are rational, and our judgments about their rationality tends to influence which non-instrumental motives we actually have.

Consider, for example, two different ways in which an agent might become convinced that she should eat a vegan diet. One way is for her to start off with a strong aversion to contributing to the suffering and death of animals, and to become informed that animals in the dairy and egg industries are routinely abused and killed very early in their lives after they exceed peak productivity.[13] But a second way for is for her to start off knowing about the conditions of the animals, and to become convinced by philosophical arguments that she *should* be averse to contributing to the suffering and death of non-human animals. These might appeal, for instance, to intuitions about how one should treat space aliens with the same psychology as humans, how one should treat humans with mental lives comparable to non-human animals, and whether a being's species membership independent of her psychology should make a difference to how one treats her.[14]

While the first way of being convinced to eat a vegan diet might involve nothing more than discovering what would satisfy one's existing non-instrumental motives, the second way seems to go beyond this. It seems to involve a kind of inquiry into what one *should* be non-instrumentally motivated to do, which works by seeking out a reflective equilibrium among one's various intuitions about what kinds of things to avoid doing, what affects one's reasons to do things, general principles about what to do, and what one should do in particular cases.[15] When as a result of this process we judge that we should pursue a new end like avoiding harm to non-human animals, our judgment tends to give rise to new non-instrumental motives like aversion to harming animals for their own sakes.

We will consider three attempts to reconcile Analytic Humeanism with the kind of reflective equilibrium inquiry at work in the second way of being convinced to be vegan. We argue that all three of them fail, but we think that their failures bring out important features of normative judgments that an adequate metanormative theory must explain.

First, the Humean might contend that the agent is moved by something like a higher-order desire to have the motives that reflective equilibrium methods prescribe, together with a belief that they prescribe aversion to harming animals. However, a motivation to bring about a state can only bring it about by getting one to *do* things that bring it about.[16] This seems

[12] Even when we think we should satisfy an irrational motive in order to avoid the irksomeness or unpleasantness of its remaining unsatisfied, we take our act to serve a perfectly rational motive to avoid annoyance or unpleasantness.
[13] See for instance (Mason and Singer 1990, especially 5-6, 39-40 and 10-14).
[14] See for instance (McMahan 2002, 2003)
[15] For characterizations of normative inquiry as seeking such a reflective equilibrium, see for instance (Goodman 1954), (Rawls 1971), (Daniels 1979), and (McMahan 2000). This kind of inquiry inevitably involves discounting and debunking certain intuitions, and our characterization of "reflective equilibrium methods" is meant to include everything ranging from approaches like that advocated by Singer (1974), which are more skeptical of intuitions about particular cases, to approaches like that advocated by Kamm (1993), which are more skeptical of intuitions about general principles.
[16] That is, absent auxiliary apparatus like other people reading our minds and bringing about what they see we're motivated to bring about.

to follow simply from the functional role of motivations as states that combine with beliefs to directly produce *action*, as depicted below in Figure 1.

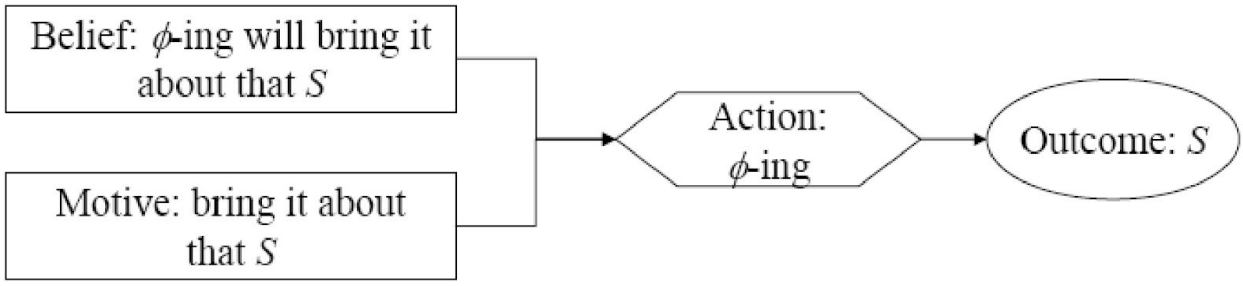

Figure 1

Hence, a motivation to have a new motive can only cause one to have it by causing one to do things to get oneself to have it. Such actions might include taking pills, classically conditioning oneself, or paying selective attention to certain things.

Thus, a desire to have the motives reflective equilibrium methods prescribe and a belief that they prescribe aversion to harming animals could only cause the aversion by causing one to do these kinds of things to get oneself to have it. But this is not how philosophical reasoning typically guides our motives. Coming to the conclusion that one should, for instance, avoid harming animals as an end in itself can *directly* cause aversion to harming them without the mediation of actions undertaken to get oneself to have this aversion. This kind of direct influence is symmetric to that of judgments about evidence on beliefs. Judging that one's evidence supports believing that there are no deities can directly cause one to believe that there are none without one's having to do anything to get oneself to believe this. But a mere desire to believe that there are deities cannot cause theistic belief without first causing one to do things to bring the belief about.

Second, the Humean might contend that in the second way of being convinced to be vegan the agent becomes aware of a strong aversion to harming non-human animals that she had all along. However, motivations cause one to do what they are *actually* motivations to do, *not* what one simply thinks they are motivations to do. One might, for instance, have a desire to approach someone sitting at a bar to whom one is attracted, which one mistakes for a desire for beer. In such a case, one will expect that if getting beer and approaching the person come apart, one will do what procures beer rather than what gets one close to the person. But if one is actually motivated to approach the person rather than procure beer, the motive will (*ceteris paribus*) cause one to violate one's expectations and do what brings one near the person rather

than what procures beer.[17] If the person leaves the bar for the sandwich shop, one may find oneself doing so as well, even though one expected that one would stay and drink.

Thus, if prior to philosophical inquiry an agent had a strong aversion to contributing to harming animals of which she was unaware, this motive would *already* have combined with her belief that non-veganism contributes to their harm and caused her to eat a vegan diet. If all reflective equilibrium inquiry did was make the agent aware of her motive, it would not change her from non-vegan to vegan; it would simply change her from a vegan with a poorer understanding of her behavior to a vegan with a better understanding of it. But reflective equilibrium inquiry can make vegans out of non-vegans. In general, when philosophical inquiry causes one to think that one should do something, it tends to supply new motivation to do it.

A final Humean response might be that in the second way of being convinced to be vegan the agent is moved by something like a first-order desire to *do* whatever reflective equilibrium methods prescribe we do. The first problem with this is that, rather than motivating the agent to do one thing or another simply as means to the end of conforming to the dictates of reflective equilibrium methods, the agent's philosophical inquiry seems to alter her non-instrumental motives themselves. Philosophical arguments for veganism of the kind we mentioned contend, for instance, that just as we should be non-instrumentally averse to harming mentally disabled humans, so too we should be averse to harming mentally comparable animals for their own sakes. Being convinced by this kind of argument tends to directly generate non-instrumental aversion to harming the animals.

To deny this intuitive picture seems to paint philosophical inquiry as necessarily producing the wrong kind of motives, or those Williams (1976) might object to as involving "one thought too many." It seems, for instance, that a philosophical argument against racism can convince someone to improve her treatment of members of other races. But something would surely be amiss it convinced her treat them better simply as a means of doing whatever philosophical methods prescribe. Surely the case against racism purports to show that the person

[17] We should stress that this causal propensity will determine what one does when all else is held equal. One can of course have countervailing motives that prevent one's procuring beer. One can also have motives like those to avoid the frustration of unfulfilled desires, which will cause different actions depending upon one's views about what one desires. One can similarly take evidence about one's desires as evidence about what one would enjoy, which, in combination with motives to do what one would enjoy, can cause action in a way that depends upon one's views about what one desires. These are not, however, instances of action produced by the mere combination of beliefs that one is motivated to bring something about and beliefs that a certain action will bring it about. Rather, in these cases motives to avoid frustration or procure enjoyment *themselves* (and *not* beliefs about them) combine with beliefs about what will bring about *their* satisfaction to produce action.

It also seems plausible that there is a mechanism that causes some of our motives to conform to our theories or narratives about the kinds of motives we have. In some cases this might generate (or strengthen) motives like desires for beer and eliminate (or weaken) motives like desires to approach a person to whom one is attracted. We think that this process is the result of our accepting norms that prescribe something like having those motives that would endow the true explanation of our motives with epistemic virtues (like simplicity, unity, and so on), and we suspect that this is much of what is correct about accounts of practical reason like that of Velleman (2000) and perhaps also Korsgaard (1996). As we will argue below, the acceptance of a norm for motivation is a distinct kind of mental state that can, unlike a desire to have a motive, directly influence the motives we have.

should improve her treatment of members of other races for their own sakes, not as just as raw materials for the enactment of the dictates of reflective equilibrium methods![18]

A second problem with this Analytic Humean response is that we use reflective equilibrium methods to try to settle the whole of the basic question of what to do, not just how one consideration bears on it. But according to this response, settling what reflective equilibrium methods prescribe would only settle what would satisfy one motive among many. Since Analytic Humeans maintain that we have reason to satisfy each of our non-instrumental motives, their responding in this way seems to commit them to the view that settling what reflective equilibrium methods prescribe settles only one question among many that bear on what to do.

4. Deliberation and Norms

We have seen that in contrast to states like desires to have beliefs and desires, or beliefs about desires and beliefs, normative judgments can exert a direct, non-behavior-mediated effect on both non-instrumental motives and beliefs. This suggests that just as beliefs and motives have the functional role of combining to directly cause action, so too there is another kind of mental state at work that has the functional role of revising beliefs and motives themselves. It is natural to speak of the mental states that play this role as states of the acceptance of norms for belief and motivation. We depict the relationships between the functional roles of beliefs, motives, and accepted norms below in figure 2:

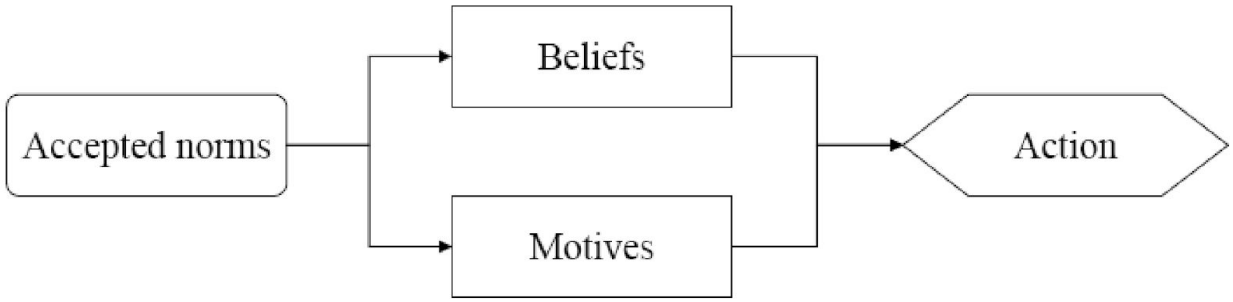

Figure 2

While the norms we accept exert causal influence on our beliefs and motives, there are several important ways in which we can fail to conform to their prescriptions. First, there are cases in which a norm we accept prescribes a response but we fail to recognize that it does so. We might, for instance, believe both:

Q: Quantum mechanics is true, and

[18] A similar Humean response to that we are considering would be to claim that the agent is moved by a *de dicto* desire to avoid doing whatever is morally wrong. Smith (1994) provides an argument against this view that is analogous to that we have given against the view that the agent is motivated by a *de dicto* desire to conform to the dictates of reflective equilibrium methods. Another problem with this response is that, as we mention below, we use reflective equilibrium methods to determine the entirety of the basic question of what to do – which extends beyond determining the moral status of our conduct. For instance, Nozick's (1974) experience machine argument can strengthen our motivations to pursue things like genuine achievement and knowledge for their own sakes, not just as means of doing what reflective equilibrium methods tell us to do.

M: If Quantum mechanics is true, then there are many worlds.

Although we accept *modus ponens*, we may fail to infer that there are many worlds from *Q* and *M* by simply overlooking *ponens*' applicability to our situation – say because we fail to "put *Q* and *M* together."

Second, there are cases in which we recognize that a norm we accept prescribes a response that we nevertheless fail to have. We might, for instance, believe *Q* and *M*, and recognize that we should infer from them that there are many worlds, but simply find ourselves unable to believe that there are such worlds. When this happens we tend to experience a species of what psychologists refer to as "cognitive dissonance."[19] This dissonance is distinctive, however, in that it does not feel as though we are torn or of two minds about an issue. If anything, it feels rather like we are weak, deficient, or inadequate to the demands of reason. As this dissonance arises when our responses recalcitrantly fail to conform to what we think they should be, we shall call it 'recalcitrance dissonance'.

Third, there are cases in which we are caused to have a response in much the same way as when we recognize that our norms prescribe it, but where the response is not in fact prescribed by norms we accept. Consider cases of fallacious deductive inference. We might, for instance, believe that:

C: If society should be Communist, we should redistribute income,

F: Society shouldn't be Communist,

and infer from *C* and *F* that:

R: We shouldn't redistribute income.

Although we can make inferences like this, it certainly does not seem that we actually accept norms that prescribe denying the antecedent. In the case of such inferences, it seems that a mere appearance that *C* and *F* commit us to believing *R* causes us to believe *R*. But the direct influence of such erroneous appearances of commitment seems identical to that of appearances that correspond to the prescriptions of norms we accept, like that to the effect that *Q* and *M* commit us to belief in many worlds. Indeed, if it looks to us like *C* and *F* commit us to *R* and we fail to believe *R*, we will tend to have the same kind of recalcitrance dissonance that we have when we fail to believe in many worlds despite its looking like *Q* and *M* commit us to such belief.

The fact that we can fail to conform to the norms we accept in the first and third ways just described suggests that our attitudes are governed by *representations* of what the norms we accept prescribe. These representations can be mistaken, and when they are we have a tendency to conform to what we represent our norms as prescribing rather than what they actually prescribe. If this is right, then we must have a way of representing the norms we accept without being able to know all of their prescriptions. At the same time, the representations of our norms that play a role in inference surely do not have a mode of presentation like THE NORMS I ACCEPT, WHATEVER THEY ARE. Perhaps the mental states that represent what our norms prescribe do so in virtue of bearing a certain nomic relation to the mental states that constitute our acceptance of these norms. Candidates for this nomic relation might resemble the kind of information carrying under ideal conditions discussed by Dretske (1981) or the kind of asymmetric causal dependencies discussed by Fodor (1987, 1990).

[19] See for instance (Festinger 1957).

Interestingly, a significant and rather diverse group of philosophers have been attracted to the idea that we use reflective equilibrium methods of normative inquiry to discover our own deepest commitments. The idea is that such methods seek to uncover the structure of an underlying "practice," "capacity," "sense," or "set of values" that generates our normative intuitions and judgments, but to which we lack conscious access.[20] Such methods can, however, cause us to extend and revise our normative views in significant and even radical ways. Many kinds of distorting influences, including wishful thinking, emotional biases, and faulty theorizing can cause our normative intuitions and judgments to fail to reflect our underlying commitments. In seeking a unification of our normative intuitions we attempt to determine which can be debunked as products of distortion and construct a theory of our commitments that best explains the intuition they generate. In this way we achieve access to our genuine commitments in much the way empirical inquiry achieves access to the external world by constructing a best explanation of our perceptual experiences, which are causally sensitive but by no means infallible guides to it.

Now, since our normative judgments directly guide our attitudes in a way unlike representations of beliefs and desires, it seems that the reflective equilibrium methods that generate these judgments cannot be attempts to discover what we believe or desire. But what if, rather than beliefs or desires, the underlying "sense" or "values" we use reflective equilibrium methods to uncover are the norms we accept? As we have seen, the norms we accept generate representations of what they prescribe, but other causal factors can cause these representations to deviate from our norms' genuine prescriptions. Quite independently of their veracity, these representations that our norms prescribe a response exert direct causal influence on our coming to have it, and should this influence fail to determine our response, we will tend to experience recalcitrance dissonance. But reflective equilibrium methods can correct and extend these attitude-guiding representations of our norms' prescriptions, providing us with an *a priori* form of access to synthetic facts about what the norms we accept prescribe.[21]

It would seem, then, that the identity of normative inquiry with inquiry into what the norms we accept prescribe could explain normative inquiry's causal and epistemic features. In particular, the identity of our deliberations about what ends to pursue with inquiries into what non-instrumental motives our norms prescribe would explain how such deliberations directly influence our motives and can achieve *a priori* access to a synthetic subject matter.

---

[20] See for instance (Goodman 1954), (Rawls 1971), (M.B.E. Smith 1977, 1979), (Fischer and Ravizza 1992), (Kamm 1993), (Unger 1996), and (McMahan 2000).

[21] This form of access would rather straightforwardly count as *a priori* in the more liberal senses discussed by (Boghossian and Peacocke 2000), according to which *a priori* access is access independent of sensory experience. But the access afforded by reflective equilibrium methods to the prescriptions of the norms we accept may qualify as *a priori* in a stronger sense too, since it is a from of access to rather general facts that are at least in part about abstracta, and – if we are correct about what rational intuitions and insights are – it is a form of access by means of "pure reason" or cognitions of this sort alone.

It might be objected that reflective equilibrium methods are an *a posteriori* form of access to our norms because they involve debunking-arguments that pursue empirical hypotheses about the actual origins of certain intuitions. But it seems that a form of evidence (like normative intuition) can be capable of being defeated or strengthened by empirical evidence without losing its status as *a priori* in an interesting sense (see for instance (Russell 2007) and (Bonjour 1998)).

As attractive as these identifications might thus be, some philosophers have balked at identifying our more critical normative inquiries with attempts to determine our own underlying commitments.[22] The crux of the worry seems to be that we can evaluate the rationality of our own commitments, or the norms we accept, themselves. We might start out accepting norms, like *Desire only pleasure for its own sake!*, or *Feel more averse to harming members of your own race!*, that philosophical arguments convince us to reject.

It is crucial to note, however, that these philosophical arguments work by means of the same reflective equilibrium methods we have been discussing, and seem to have the same causal and epistemic features. Coming to think that one should accept or reject a norm exerts direct causal influence on one's accepting or rejecting it, and the indecisiveness of this influence tends to engender recalcitrance dissonance. Deliberation about what norms to accept is a kind of *a priori* inquiry that seems capable of hooking onto a synthetic subject matter by debunking some and seeking out a best explanation of others of our intuitions about what to accept.

As such, the general considerations that support identifying the normative evaluation of a response with holding it up to norms we accept equally support identifying the normative evaluation of a norm we accept with evaluating it against higher-order norms we accept. Someone might, for instance, accept norms prescribing greater concern for her own race because they appear to be licensed by higher-order norms that prescribe greater concern for those to whom she is specially related. But this appearance may be due to a conflation of membership in the same race with somewhat correlated features like degree of personal contact. The racist may discover that her norms about personal relations privilege only the latter by consulting her intuitions about hypothetical cases and the relevance of what race-membership actually comes to.

To accept a norm that prescribes response *R* is roughly to be in a state which is such that representations of it cause responses of *R*'s kind on pain of recalcitrance dissonance, and accurate representations of it cause response *R*. To accept a "first-order" norm for belief or motivation is to be in a state that in this way regulates beliefs and motives without regulating any intermediary states of norm acceptance. To accept an "n+1st-order" norm ($n \geq 1$) is to be in a state that regulates the acceptance of n^{th}-order norms. By regulating the acceptance of lower-order norms, the higher-order norms we accept actually govern responses of the kind prescribed by these lower order norms, and constitute norms for these responses as well.[23] What we have seen, then, is that the causal and epistemic features of deliberation about whether to accept and have the responses prescribed by lower-order norms can be explained by identifying it with inquiry into whether the lower-order norms' prescriptions are seconded by higher-order norms. Judging that one should respond as a lower-order norm prescribes thus seems contingent upon an appearance that the response is ultimately prescribed by higher-order norms that endorse the lower-order norm.

Now, at some point our psychologies will run out of norms against which to assess other norms.[24] We will terminate at some highest-order or most fundamental norms that govern our acceptance or rejection of all lower order norms and their prescriptions. Since norms have

[22] See for instance (Rawls 1971, 1974) and (Daniels 1979, 1980).

[23] Thus, to keep n^{th}-order norms from themselves counting as n-k^{th}-order norms (for n-$k > 1$), we should understand n+1st-order norm acceptance as a state that regulates n^{th}-order norm acceptance and the acceptance norms of no higher orders.

[24] Which will be the case, moreover, for any entity with finite psychological capacities.

prescriptions for the responses they regulate, the fundamental norms that regulate our acceptance of norms for belief and motivation will constitute our most fundamental norms for beliefs, motives, and the actions they motivate.

If this account of normative inquiry is correct, we are in no position to say exactly what our fundamental norms are until we know the true general theory of what to believe and pursue. We suspect that our most fundamental norms for belief include deductive norms like *modus ponens* and ampliative norms like inference to the best explanation.[25] Our coming to accept these norms may have been an evolutionary adaptation that enabled our ancestors to form accurate beliefs under a wide range of novel and complex conditions.[26] Similar selection pressures may have caused us to accept fundamental norms for motivation that enabled our various motivational systems to play their adaptive roles more flexibly and in ways better suited to novel and complicated environments. We suspect that these norms are rather complex. But proponents of different theories about what we have reason to do will have their own views about these fundamental norms. A utilitarian about rationality might, for instance, contend that we all fundamentally accept the principle of utility and only tend to judge that utilitarianism is false because we mistake things that are typically optimific for things our norms prescribe doing for their own sakes.

As we have seen, our questions about what to do are questions about what will serve ends that are worth pursuing for their own sake. We use reflective equilibrium methods to determine which ends to pursue and combine them with our beliefs about what will achieve these ends to determine what to do. We have also seen that identifying inquiry into what to pursue with inquiry into the prescriptions of our most fundamental norms for motivation can explain its influence on motivation and its ability to hook onto facts about its subject matter *a priori*. It would seem, then, that we can explain the central causal and epistemic features of inquiry into what to do by identifying it with inquiry into which actions are prescribed by the most fundamental norms we accept.

If inquiry into what to do can be explained in this way, it might seem that questions and judgments about what to do are a particular mode of presentation of questions and judgments about what our most fundamental norms prescribe - the mode of presentation we encounter in deliberation, perhaps in virtue of these states bearing the right nomic relation to the states that constitute our acceptance of these norms. We shall call this view about judgments concerning one's own reasons for action:

First-Person Norm Descriptivism:
To judge that one has reason to φ is to believe under a deliberative mode of presentation that one's most fundamental norms prescribe that one φ.

One might, however, attempt to resist the inference from our explanation of normative inquiry to First-Person Norm Descriptivism if one thought that our normative judgments "point through" the norms we accept to something else. Much as representations of magnified images can be about the microscopic objects that caused them, the idea might be that the normative judgments are about something in the external world that caused us to accept the norms we do.

[25] Where determining what it is for something to be a "best explanation" is a serious task of normative epistemology.
[26] For an excellent and engaging discussion of these environmental conditions, see (Quartz and Sejnowski 2002).

The problem with this is that the causal genesis of our fundamental norms seems irrelevant to the content of our normative judgments. Considerations of parsimony dictate that no *sui generis* normative facts explain our acceptance of norms. Suppose it turned out that we accept the fundamental norms we do because of some particular evolutionary story, or because of a certain divine creation, or because of the details of how we sprang into existence a few moments ago as a result of lighting hitting a swamp. None of this seems to make a difference to what we are thinking about when we make judgments about what to do.

When facts about the genesis of certain norms matters for figuring out whether to accept them, these facts bear on whether the norms fall short of an independent standard that guides our acceptance or rejection. But since our fundamental norms are the ultimate standards by which we assess our norms, their origins cannot matter in this way. The normative judgments that guide us seem to point as far as the prescriptions of our fundamental norms but no further.

5. Norm Descriptivism vs. Relativism and Expressivism

We have thus argued that First-Person Norm Descriptivism is the best explanation of the causal and epistemic features of judgments about our own reasons for action. Now, when we make judgments about another agent's reasons, we make judgments about the same things she does. The judgment we express with 'you shouldn't φ' contradicts her judgment that she should. The agent seems able to assess the truth of our judgment using the same reflective equilibrium methods she uses to assess her own, and coming to believe that our judgment is true will directly guide her responses in the same way as her own. Our thoughts about her reasons thus seem to be thoughts to the effect that the answers she seeks in deliberation are thus and so.

If this is right, then the considerations in favor of First-Person Norm Descriptivism suggest that our judgments about another agent's reasons are judgments about what her most fundamental norms prescribe under a mode of presentation that we fix by reference to her deliberations. This yields the following general view of judgments about reasons for action:

Global Norm Descriptivism:
To judge that agent *A* has reason to φ is to believe under a mode of presentation derived from *A*'s deliberations that the most fundamental norms *A* accepts prescribe that *A* φ.

One might worry that Global Norm Descriptivism (hereafter just 'Norm Descriptivism') fails to capture the apparent platitude that to judge that another agent has reason to do something is to endorse her doing it, or to think that one would have reason to do it oneself if one were in her circumstances. Of course, Norm Descriptivism easily captures this platitude if "an agent's circumstances" are allowed to include facts about the norms she most fundamentally accepts. But the kind of circumstances intended by the platitude are presumably those which authoritative norms have prescriptions for rather than those which determine which norms are authoritative. The alleged platitude thus seems to amount to the idea that that the same basic set of norms is authoritative for each agent, or prescribes what each agent should in fact do.

Now, as we mentioned, we suspect that we came to accept the fundamental norms we do as a universal human adaptation to variable and complex environments. Our experience with shared normative inquiry also suggests that a shared set of fundamental norms is responsible for

our normative intuitions. There is far more similarity in people's *intuitions* about cases, relevant differences, and principles (once fully clarified and understood) than there is among their views about how to account for them. Even where specific intuitions differ, there seems to be substantial similarity in surrounding intuitions, including about such epistemic considerations as what kinds of origins of intuitions look suspicious and what features look like theoretical problems for intuitions.

This overlap in intuitions and the problems and prospects of accounting for them is what seems to make shared normative inquiry possible. Against a background of shared fundamental norms that generate such overlap, we can speak simply of 'the thing to do in a circumstance', which will be the same for all of us. Perhaps, then, we tend to think that the same norms are authoritative for everyone because our shared normative inquiry presupposes that we all accept the same fundamental norms, and this presupposition is pretty much correct. If this is right, then Norm Descriptivism and the fact that we all accept the same fundamental norms can explain how the same norms are authoritative for everyone and capture the platitude about endorsement.

Of course, if it is a fact that all agents accept the same fundamental norms, it seems to be a contingent fact, and unlikely to be a fact at all if we take its universal quantification to be literally unrestricted. Mutations, quirky developmental trajectories, and injuries make it almost inevitable that some humans will fail to have traits that were universal human adaptations. One could try saying things like "we rigidly fix the reference of 'agent' as *being that accepts the fundamental norms we actually do!*" But this merely distracts attention from more interesting questions about what correctly answers the action guiding questions asked by beings that seem able to figure out such answers by reflective equilibrium methods.

If First-Person Norm Descriptivism is right, then a respect in which a being's fundamental norms for action differ from ours is a respect in which her deliberations about what to do aim at something different from ours. It is a respect in which perfectly careful inquiry from intuitions that accurately represent their subject matter would lead each of us to think we should perform different actions. It seems quite natural to think that in such cases both of us would be correct, and that we simply have reason to do different things. Of course, if the other agent has reason to do something that we have serious reason to prevent, we might have very good reason not to let her know this, and even to lie to her and trick her about her reasons if need be. But conflicts of interest and reasons not to let an agent think she has reason to do something no more signal the presence of disagreement in normative judgment here than they do in competitive games.

Some people might, however, accept our argument for First-Person Norm Descriptivism but think for some reason that it is very important for us to be able to truly or sincerely say of agents that accept different fundamental norms that they have reason to do what we do. These people might be reluctant to abandon pre-theoretical intuitions to the effect that we have the same reasons. To retain these intuitions, they might try to reduce our judgments about what other agents should do to judgments from our own deliberative perspective about what to do in their circumstances.

These opponents of Global Norm Descriptivism may also think that there is far less similarity than we do in the fundamental norms people accept, and find themselves reluctant to think that the savvy participant in normative discussions must go in for so much trickery. If so,

they will need an account of how sincere expressions of normative judgments made from deliberative perspectives with different targets can seem like genuine disagreements rather than mere differences in attitude. A sketch of such an account goes back at least to Stevenson (1937), and Gibbard (1990) develops a rich and impressive version of it. In our context the idea would be that there is a psychic mechanism by means of which simply expressing our judgments about (and perhaps our acceptance of) our fundamental norms tends to cause our audience to accept them as well. This would support the following versions of Norm Relativism or Expressivism:

Norm Relativism (Expressivism):
To judge that agent *A* has reason to *φ* is to (accept one's most fundamental norms and) believe under a deliberative mode of presentation that one's most fundamental norms prescribe *φ*-ing in *A*'s circumstances.[27]

A first problem with these views is that their attempts to retain our ability to truly or sincerely judge that agents with different fundamental norms should do what we should do end up portraying all normative discussion as dishonest and akin to brainwashing. Telling another agent that she should do something purports to tell her something that, just like her own judgments about reasons, is true just in case it correctly answers her deliberative questions about what to do. But on these accounts one's judgments about her reasons are not what they purport to be - their truth is in no way dependent upon their ability to correctly answer her deliberative questions about what her fundamental norms prescribe. Like brainwashing, the influence of interpersonal normative discussion on an agent's attitudes and behavior is independent of the influence exerted by her reasoning about what to do and the facts that her reasoning seeks out.

A second problem with these kinds of views is that, as Egan (2006) has pointed out, they entail that each agent can arrogantly claim to be immune to a kind of normative error to which others are prone. These versions of Norm Relativism and Expressivism subscribe to our account of access to one's own reasons in terms of the identity of the deliberative question about what to do with the question of what one's fundamental norms prescribe. They entail that if one manages to accurately discern the prescriptions of one's most fundamental norms, one is guaranteed to be correct about what one has reason to do. But these versions of relativism and expressivism hold that other agents are correct about what they should do only insofar as they think that they should do what one's own norms prescribe. Since other agents may accept fundamental norms that prescribe doing otherwise, these views entail that they might

[27] This version of Norm Expressivism should not be equated with the view by the same name that Gibbard (1990) develops and defends. This is because we understand the state of norm acceptance rather differently than Gibbard. What Gibbard calls 'norm acceptance' is much closer to the states we identify as fallible representations of what the norms we accept prescribe, in that they (rather than representations of them) exert direct causal influence on attitudes and play roles akin to the generation of recalcitrance should this influence fail to be decisive.

Because it conflicts with our account of first-person deliberation, Gibbard's version of Norm Expressivism seems to escape the problems we develop for the versions of Norm Relativism and Expressivism we discuss. Our primary argument here against Gibbard's Norm Expressivism is our argument that First-Person Norm Descriptivism best explains the causal and epistemic features of first-person normative inquiry. While it is largely a story for another time, we touch briefly on our worries about the ability of Gibbard's kind of account to tell a plausible story about first-person normative inquiry in our response to Moore-like challenges below. We should also mention that the coordinative evolutionary story Gibbard tells about how avowals of the norms we accept fundamentally influence others is far more plausible for moral norms than for other norms, like those for belief and prudent action.

successfully discern what their fundamental norms prescribe and yet be wrong about what they should do.

Thus, according to these versions of Norm Relativism and Expressivism, other agents can deliberate ever so carefully from intuitions that accurately track their content, yet fail to get things right about what to do. They are vulnerable to a kind of "normative blindness" – their fundamental norms and the intuitions that reflect them might simply fail to correspond to something like an independent normative reality. But since the standards of this independent normative reality are set by one's own fundamental norms, it is inconceivable that one could suffer from the same fundamental blindness as other agents. Surely this seems wrong.

We think that these costs of attempts to retain intuitions that agents with different fundamental norms have the same reasons we do outweigh their benefits. They seem to run afoul of the facts that the dynamics of normative discussion mirror those of first-person deliberation and that other agents have the same kind of access to their reasons that we do to ours. Norm Descriptivism's identification of judgments about another agent's reasons with judgments about the target of her deliberations seems to provide a far better explanation of these features of normative thought. Moreover, there is reason to think that interpersonal normative inquiry presupposes a background of shared fundamental norms, and the general truth of this presupposition is suggested by our experience with normative inquiry and evolutionary considerations. If this is right, it would not be surprising for us to mistake reasons shared by almost all human agents for reasons shared by all conceivable agents.

6. Replies to Objections

We have argued, then, that Norm Descriptivism provides the best explanation of the causal and epistemic features of both first- and third-person judgments about reasons. We turn now to defending Norm Descriptivism against some objections.

First, one might worry that Norm Descriptivism cannot explain the normative force of morality. To the contrary, we think that Norm Descriptivism can be combined with an independently plausible account of moral reasons to vindicate them. We think that analyses like the following best capture the content and normative force of moral judgments:

The Fitting Attitude Analysis of Moral Wrongness:
Agent *A*'s act of φ-ing is morally wrong if and only if *A* should feel obligated not to φ

where to feel obligated not to do something is to have a kind of prospective guilt-tinged aversion to doing it.[28] This kind of analysis can explain the intuitively wide variety of acts that can

[28] This is the attitude that one characteristically feels upon contemplating the prospect of doing something that one takes to be morally wrong. It is discussed admirably by Brandt (1959, 117-118), and was what Mill (1863) described as an "internal sanction of duty…a feeling in our own mind… attendant on violation of duty, which in properly cultivated moral natures rises, in the more serious cases, into shrinking from it as an impossibility," and "a mass of feeling which must be broken through in order to do what violates our standard of right."

Some might contend that this attitude involves a subconscious judgment about moral obligation or wrongness. It seems, however, that one can feel this attitude towards doing something and judge that one is not

coherently (if quite falsely) be judged morally wrong. It can also explain how moral judgments have the causal and epistemic properties of normative judgments by subsuming them under the general phenomenon of judgments about reasons for attitudes. We think that similar analyses can be given of judgments about moral blameworthiness, estimability, and disestimability in terms of reasons to feel guilt and anger, moral esteem, and moral disesteem.[29]

Moral emotions like these are motivational states – feeling obligated not to do something involves motivation not to do it, esteeming an act involves motivation to emulate it, and so on. According to Norm Descriptivism, judgments that an agent should have motives like these are judgments that her fundamental norms prescribe having them, which entail that her norms prescribe performing the actions these motives motivate. In conjunction with the above analyses of moral judgments, Norm Descriptivism can understand judgments that an agent has moral reason to do something as judgments that her fundamental norms prescribe doing it out of moral emotions.[30]

Now, this understanding of moral reasons would entail that agents who fundamentally accept no norms for moral emotion lack these reasons. In fact, there is some evidence that certain severely sociopathic or psychopathic people may be psychologically incapable of moral emotions, as well as attitudes like care and shame.[31] If their minds really are set up in such a way that they are incapable of these attitudes, then their fundamental norms cannot prescribe them (for this would require causal propensities for them to have such emotions). Our understanding of moral reasons would thus entail that these sociopaths lack moral reasons, and that their actions are never morally wrong.

It is of course consistent with this that any horrible things these agents do are still bad and such that we (non-sociopaths) have reason to prevent them from doing them. We are quite used to the idea that horrible things done by many beings are bad and to be prevented without their being in any way wrongful. This is surely our attitude towards such things as sharks attacking our friends and coyotes attacking our companion animals.

We do think that sociopaths are different from sharks and coyotes in that they are agents who are subject to some reasons, like reasons for belief and prudent action. We think, however, that the best explanation of why sharks and coyotes cannot wrongfully harm is that they cannot reason their way to refraining from inflicting harm out of feelings of obligation. This explanation equally entails that sociopaths cannot wrongfully harm if they cannot reason their way to moral emotions. The fact that they can reason their way to other attitudes is irrelevant.

---

obligated to do it without any conflict in judgment (or indeed any conflict between judgment and intuition). But a full discussion of these issues is beyond the scope of this paper.

[29]Cf. Gibbard's (1990, 40-45, 126-127) analysis of MORAL BLAMEWORTHINESS and Brandt's (1946, 113) general suggestion that "'X is Y-able'…means that 'X is a fitting object of Y-attitude (or emotion).'"

[30] One clarification is in order. For reasons we have discussed, we suspect that pretty much all human agents who accept moral norms accept the same fundamental ones. But suppose there were space-aliens who accepted fundamental norms that prescribed that they feel obligated to torture us as an end in itself. Surely it would be false to say that it would be morally wrong for these agents to fail to torture us. Norm Descriptivism can explain this in terms of the connection between moral wrongness and moral blameworthiness. Judgments that an act is morally wrong entail that it is blameworthy absent excuses like diminished responsibility. But to judge an act blameworthy involves judging that a contextually determined set of agents that includes oneself have reason to be angry at its author. Since we lack reason to be angry at fully responsible aliens for failing to torture us, it would be false for us to judge that their failures to torture us are wrongful, even if they do have reason to feel obligated to torture us.

[31] See for instance (Mealey 1995).

A second objection to Norm Descriptivism might be reminiscent of Moore's "open question argument." It looks like Norm Descriptivism gives us an analytic identity between facts about what one should do and facts about the prescriptions of one's fundamental norms. But it certainly seems that one can ask, "Should I really continue to accept and follow the prescriptions of my fundamental norms?" How does Norm Descriptivism account for this?

If one's question about whether to accept and follow one's fundamental norms is a genuine question about what to accept and do, then answers to it must directly guide what one accepts and does in the ways characteristic of normative judgments. It thus seems that one cannot be asking such things as whether one's continuing to accept the system of norms one does will make one happy, satisfy one's desires, be approved of in one's society, or conform to rules that one does not currently accept.

Some might contend that such questions about what to accept are not about anything descriptive at all. We find it rather obscure what they would then be. Perhaps they are supposed to express requests for or signals of receptiveness to the kind of influence that relativists and expressivists might posit to make sense of normative discussion.[32] But questions about what to accept do not request any old kind of influential utterances; they request correct answers to what they are asking. The search for such correct answers seems to aim at a state of knowledge in which the truth of one's beliefs about the answers helps explain why one holds them. But this kind of explanatory work can only be done by descriptive facts.

We think that the only way to explain the causal and epistemic features of questions about whether to accept and follow one's fundamental norms is to identify it with an evaluation of these norms against themselves. This is what Norm Descriptivism does by interpreting it as a question of whether one's fundamental norms, conceived under a deliberative mode of presentation, prescribe accepting these norms, conceived of as THE FUNDAMENTAL NORMS I ACCEPT. Since the deliberative mode of presentation does not refer to one's norms under the latter description, it is by no means obvious to the agent that her question is an assessment of her fundamental norms in terms of themselves.[33]

In fact, it seems possible to conceive of agents who accept fundamental norms that prescribe rejecting themselves or their own prescriptions. In wondering whether one should

---

[32] If there are no actual others with whom one is talking, perhaps one makes this request or signals this receptiveness to influence to the voices in one's head or the imaginary interlocutors with whom one rehearses for normative discussions. On this notion of deliberation as rehearsal for normative discussion, see (Gibbard 1990, especially 74-75, 81). << The suggestion that claims of normative uncertainty request or signal receptiveness to influence is also Gibbard's <<(personal correspondence)>>.

[33] Although Norm Descriptivism identifies normative judgments with representations of agents' fundamental norms under modes of presentation that do not describe them, it still maintains an analytic identity between facts about agents' reasons and facts their norms. According to the view, it is an analytic truth that a state would not count as a normative judgment if it failed to play the right kind of roles in deliberation and attitude guidance, which involve representing one's fundamental norms under a deliberative mode of presentation (perhaps by bearing the right kind of nomic relation to states of norm acceptance).

We believe that a similar kind of analytic identity is maintained by analytic functionalists about qualia who think that our ordinary qualia concepts are phenomenal. According to these views, it's analytic that whatever states play the qualia-roles are qualia, even though we ordinarily represent these states with phenomenal concepts that do not describe the qualia-roles. The analyticity is maintained by the contention that it is an analytic truth that a state that failed to bear the right kind of representational relation to whatever states play the qualia roles would not count as a (phenomenal) qualia concept.

accept and follow the prescriptions of one's fundamental norms, one can be said to be wondering whether one's fundamental norms are self-undermining in this kind of way. Since there seems to be no indication from our reflective equilibrium methods that we are committed to self-undermining principles with no deeper, non-self-undermining principles to fall back on, we seem to have little to fear from this possibility. But one's acceptance of self-undermining fundamental norms is still quite a genuine possibility to be concerned about and to wonder about *de re* when one wonders whether to accept and follow one's own fundamental norms.

A final objection to Norm Descriptivism might be a third-person analogue of the second. It might still seem at least coherent to think that another agent's fundamental norms are simply irrational or crazy, or that she has reason not to do certain things, like torture innocents just for fun, whether or not her fundamental norms prescribe against them. We think, however, that these appearances of coherence stem from a lack of appreciation of what distinguishes an agent's having reason to do something from other normative phenomena. Return to the distinction between:

(a) an entity's doing something the occurrence of which is bad, or something we should hope it doesn't do and oppose if we can, and

(b) an entity's doing something that it has reason not to do.

The occurrence of natural disasters and attacks by sharks and coyotes are instances of (a) but not (b), while the sub-optimal play our opponents in fair competitions are instances of (b) but not (a).

We should thus ask: why, in any given case, should we think that someone's doing something like torturing innocents for fun is not only something we should hope she doesn't do and prevent her from doing, but moreover something she has reason not to do? Is it not that, above and beyond our having reason to oppose her action, we think that she could correctly reason her way to refraining from performing it? Correct reasoning, however, is not just any process by which an entity comes to do what we might think we should do in its circumstances. Were we to neurally alter a shark so that he no longer attacks innocents, he would not thereby have correctly reasoned himself to refraining from doing so. Correct reasoning seems rather to be a matter of going from what one accepts to what is genuinely prescribed by what one accepts. But if this is right, how could we maintain that an agent has reason to do something even though her fundamental norms do not prescribe doing it?

The dependence of reasons on the possibility of correct reasoning seems important for explaining why it makes no sense to think that reasons apply to the responses of entities like volcanoes, infants, and sharks. If someone were to think simultaneously that volcanoes are as we take them to be (with no mental life at all) and yet that they have reason not to, say, erupt, he would seem rather straightforwardly incoherent.

To be sure, beings like infants and sharks are unlike volcanoes in that they are capable of well being or welfare; of things literally going better or worse for them. But a response's being good or bad for a being is distinct from her having reason to have it. As Darwall (2002) has convincingly argued, judgments about a being's welfare are judgments about our reasons to want things for her out of care for her, which in no way requires her to be subject to reasons herself.

To think beings like infants and sharks not only capable of welfare but also subject to reasons, we would seem to have to think them candidates for the kind of rational criticism involved in calling someone an idiot for making a foolish decision. Infants and sharks are of

course capable of greater or lesser intelligence or learning ability, but it seems incoherent to hold them genuinely rationally criticizable on the assumption that their minds are as we take them to be. It is true that rationally criticizablility involves more than simply failing to respond to the reasons one has; an agent can fail to do this but be rationally exculpated on account of diminished responsibility. But the way beings like infants and sharks lack rational responsibility for their behavior is importantly unlike that of, say, an otherwise psychologically typical adult human whose perennially recalcitrant emotions always sway her against her better judgment. Infants and sharks are incapable of this very kind of "better judgment" that has the peculiar causal and epistemic properties of judgments about reasons. But it seems that one must be able to correctly reason one's way to these kinds of judgments in order to be subject to reasons at all.

Norm Descriptivism provides us with a straightforward explanation of why having reason to do something requires being able to do it as a result of judging that one has such reasons. According to this view, having reason to do something is a matter of accepting fundamental norms that prescribe doing it. But part of what it is to accept such norms is for the states that constitute representations of them under a deliberative mode of presentation to directly influence motivation and action accordingly. Moreover, one cannot have representations of one's norms under this mode of presentation unless one accepts norms of which they are representations.[34] So according to Norm Descriptivism, the fact that beings like infants and sharks accept no norms entails both that they lack reasons to do things and that they cannot judge that they have them.

But proponents of the independence of an agent's reasons from the fundamental norms she accepts seem unable to explain why an entity's having reasons to do things is dependent upon her ability to judge that she has them and why it is incoherent to attribute reasons to entities like infants, sharks, and volcanoes. If an entity's having reason to do something is a property that is analytically independent of its psychological states – like the property of maximizing happiness by so acting - why does it seem incoherent to attribute this property to entities that cannot represent and respond to it, like infants, sharks, and volcanoes? One might, of course, just stipulate that the property of having reason to do something is the possession of a certain non-psychological property (like maximizing happiness by so acting) and being such that one can judge that one has it. But this accommodates the phenomenon without explaining it, and in fact makes normative judgments dependent upon psychology in a much less principled way than Norm Descriptivism.

If one can accept norms for the circumstances of beings like infants and sharks who cannot make judgments about their reasons, Norm Expressivism and Relativism entail that one can coherently judge that they are subject to reasons. The Expressivist or relativist might try to prevent this by requiring, for instance, that to accept norms that prescribe having a certain response in a circumstance, one must take that circumstance to be such that one can accept norms in it, or norms that prescribe having or not having that response.[35] But this looks *ad hoc*. Why can one accept norms that prescribe having a response in a circumstance only if one could accept norms (for or against the response) in that circumstance? This does not seem to fall out of an account of norm acceptance. It rather looks tacked-on by the Expressivist or Relativist to accommodate intuitions about the incoherence of certain judgments that her view cannot explain.

[34] Because (we suspect) these states have this content in virtue of bearing the right nomic relationship to states of norm acceptance.

[35]<<We are grateful to Allan Gibbard for these suggestions.>>

We think, then, that Norm Descriptivism provides us with a better explanation than its opponents of the incoherence of holding entities like infants, sharks, and volcanoes subject to normative reasons. We think that this is important evidence that Norm Descriptivism has the best explanation of the distinction between merely judging that we should oppose an entity's doing something and judging that it has reason not to do it. If, as we suspect, Norm Descriptivism has the best explanation of this distinction, it is in a good position to debunk intuitions that it is coherent to judge that an agent has reason to do other than what her fundamental norms prescribe in terms of a lack of appreciation of what exactly her having reason to do something amounts to.

REFERENCES

Boghossian, Paul and Christopher Peacocke (eds.). 2000. Introduction. *New Essays on the A Priori*. Oxford: Clarendon Press.

BonJour, Laurence. 1998. *In Defense of Pure Reason*. Cambridge: Cambridge University Press.

Brandt, Richard B. 1946. Moral Valuation. *Ethics*, 56: 106-121.

Brandt, Richard B. 1959. *Ethical Theory*. Englewood Cliffs, N.J: Prentice Hall.

Brandt, Richard B. 1979. *A Theory of the Good and the Right*. Amherst, New York: Prometheus Books.

Daniels, Norman. 1979. Wide Reflective Equilibrium and Theory Acceptance in Ethics. *Journal of Philosophy*, 76:5: 256-82.

Daniels, Norman. 1980. On Some Methods of Ethics and Linguistics. *Philosophical Studies*, 37: 21-36.

Darwall, Stephen L. 1983. *Impartial Reason*. Ithaca: Cornell University Press.

Darwall, Stephen L. 2002. *Welfare and Rational Care*. Princeton: Princeton University Press.

Dretske, F.I. 1981. *Knowledge and the Flow of Information*. Cambridge: MIT Press.

Enoch, David. 2007. An Outline of an Argument for Robust Metanormative Realism. In Russ Shafer-Landau, ed., *Oxford Studies in Metaethics: Volume 2*. Oxford: Oxford University Press.

Egan, Andy. 2006. Quasi-Realism and Fundamental Moral Error. *Australasian Journal of Philosophy*, 85:2: 205-219.

Festinger, Leon. 1957. *A Theory of Cognitive Dissonance*. Stanford: Stanford University Press.

Firth, Roderick. 1952. Ethical Absolutism and the Ideal Observer Theory. *Philosophy and Phenomenological Research*, 12: 317-345.

Fischer, John Martin and Mark Ravizza. 1992. Ethics: Problems and Principles. Orlando: Harcourt Brace Jovanovich College Publishers.

Fodor, Jerry A. 1987. *Psychosemantics*: the Problem of Meaning in the Philosophy of Mind. Cambridge, Massachusetts: MIT Press.

Fodor, Jerry A. 1990. *A Theory of Content and Other Essays*. Cambridge, Massachusetts: MIT Press.

Gibbard, Allan. 1990. *Wise Choices, Apt Feelings*. Cambridge, Massachusetts: Harvard University Press.

Gibbard, Allan. 2003. *Thinking How to Live*. Cambridge, Massachusetts: Harvard University Press.

Goodman, Nelson. 1954. *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.

Harman, Gilbert. 1977. *The Nature of Morality: An Introduction to Ethics*. New York: Oxford University Press.

Hempel, Carl. 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science.* New York: Free Press.

Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Oxford University Press.

Kamm, Francis M. 1993. *Morality and Mortality.* Vol. 1. Oxford: Oxford University Press.

Korsgaard, Christine M. 1996. *The Sources of Normativity*. New York: Cambridge University Press.

Lewis, David. 1970. How to Define Theoretical Terms. *Journal of Philosophy*, 67: 427-446.

Mason, Jim and Peter Singer. 1990. *Animal Factories*. New York: Harmony Books.

McMahan, Jeff. 2000. Moral Intuition. In Hugh LaFollette, ed., *Blackwell Guide to Ethical Theory*. Oxford: Blackwell.

McMahan, Jeff. 2002. *The Ethics of Killing: Problems at the Margins of Life*. New York and Oxford: Oxford University Press.

McMahan, Jeff. 2003. Animals. In R.G. Frey and Christopher Wellman, eds., *Companion to Applied Ethics*. Oxford: Blackwell.

Mealey, Linda. 1995. The sociobiology of sociopathy: an integrated evolutionary model. *Behavioral and Brain Sciences* 18:523–99. Reprinted in Baron-Cohen, Simon, ed., *The Maladapted Mind*.

Mill, John Stuart. 1863. *Utilitarianism*. London: Parker, Son, and Bourn.

Nozick, Robert. 1974. *Anarchy, State, and Utopia*. United States of America: Basic Books.

Quartz, Steven R. and Terrence J. Sejnowski. 2002. *Liars, Lovers, and Heroes: What the New Brain Science Reveals About How We Become Who We Are*. New York: Harper Collins Publishers.

Rawls, John. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Rawls, John. 1974. The Independence of Moral Theory. *Proceedings and Addresses of the American Philosophical Association*, 47:5-22.

Russell, Bruce. *A Priori* Justification and Knowledge. *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/apriori/

Smith, Malcolm B.E. 1977. Rawls and Intuitionism. *Canadian Journal of Philosophy*, Supplementary Volume 3: 163-178.

Smith, Malcolm B.E. 1979. Ethical Intuitionism and Naturalism: A Reconciliation. *Canadian Journal of Philosophy*, 9: 609-629.

Smith, Michael. 1994. *The Moral Problem*. Oxford: Blackwell Publishing Ltd.

Singer, Peter. 1974. Sidgwick and Reflective Equilibrium. *Monist*, 58: 490-517.

Stevenson, Charles L. 1937. "The Emotive Meaning of Ethical Terms," *Mind* (46):14-31.

Unger, Peter. 1996. *Living High and Letting Die*. New York: Oxford University Press.

Velleman, J. David. 2000. *The Possibility of Practical Reason*. New York: Oxford University Press.

Williams, Bernard. 1976. "Persons, character, and morality." In A.O. Rorty (ed.), *The identities of persons*, Berkeley: Universtiy of California Press. Reprinted in B. Williams, *Moral Luck*, Cambridge: Cambridge University Press, 1981.