9

Direction of Fit and Motivational Cognitivism

Sergio Tenenbaum

1. Introduction

The idea of direction of fit has been found appealing by many philosophers. The idea goes back to Anscombe's famous example of the different aims of two agents: a man who is doing his shopping guided by a shopping list, and a detective compiling a list of the man's groceries as he buys them. I can't resist adding myself to the legion of writers who quote the passage from *Intention* in which she points out an important difference between the two directions of fit:

> If the list and things that the man actually buys do not agree, and if this and this alone constitutes a *mistake*, then the mistake is not in the list but in the man's performance . . . whereas if the detective's record and what the man actually buys do not agree, then the mistake is in the record.[1]

In both cases of mistake, there's a lack of fit between the world (in particular, what's in the shopping cart), and the agent's mental states (represented by what they have written in their respective lists). But the different "location" of the mistake in the two cases is supposed to show that desires and beliefs

I would like to thank Danielle Bromwich, Philip Clark, Thomas Hurka, Niko Kolodny, Jennifer Nagel, Fred Schueler, Rob Shaver, David Sobel, and two anonymous referees for Oxford University Press for insightful comments on earlier versions of the paper. Versions of this paper were read at the Ethics and Political Workshop at the University of Toronto, at the Metaethics Workshop in Madison, Wisconsin, and at 2005 Workshop on Moral Psychology at Franklin & Marshall College. I would like to thank members of the audience for a lively and very helpful discussion on all these occasions, and in particular, Jeff Seidman, who commented on the paper at the Franklin & Marshall workshop. Research on this paper has been generously supported by a grant from the Social Sciences and Humanities Research Council of Canada.

[1] Anscombe (1963: 56).

have different directions of fit; in the former case the world was supposed to fit the desires (the shopping cart should have been filled in accordance with the shopping list), whereas in the latter case the beliefs were supposed to fit the world (the detective's list should have been made in accordance with what was in the shopping cart). The example, as well as the direction of fit metaphor, seems to many to capture something very important about the different natures of belief and desire, something that might have profound implications for various views about motivation and practical reason. Perhaps the most notable among these supposed implications is the Humean Theory of Motivation, the claim, roughly, that beliefs cannot motivate on their own.[2] But, of course, philosophers haven't rested content with trying to derive these implications from one example and a metaphor. Philosophers have tried to make the notion of direction of fit more precise, or to provide various explanations of the intuition that the detective's and the shopper's attitudes are to be distinguished in terms of directions of fit.[3] There are at least two promising strategies in the literature for spelling out the notion of direction of fit. The first strategy cashes out the metaphor in terms of the different relations of counterfactual dependence between, on the one hand, belief and the world, and on the other hand desire and the world (or on some other attitude that is supposed to track the world). The second strategy appeals to existence of a constitutive relation between truth and belief: belief aims at the truth, whereas desire doesn't. I will argue that the first strategy collapses into the second. However the idea that there is a constitutive relation between belief and truth is itself rather vague, and it is hard to see how it can *explicate* the metaphor of direction of fit, rather than just replace one metaphor (the direction of fit metaphor) with another (the metaphor that belief aims at the truth). The second part of the paper examines whether we can understand the notion of direction of fit in terms of the constitutive relations that belief, but not desire, bears to the truth. I argue that there is indeed a way to cash out the metaphor of direction of fit in these terms; in particular, I argue that the metaphor is best cashed out in terms of the different formal ends guiding the inference from what I call "prima-facie" attitudes to what I call "all-out" attitudes in the theoretical and practical realms.

With this (hopefully) improved understanding of the distinct directions of fit of belief and desire, we can ask whether the fact that desire and belief have these distinct directions of fit can have the rich implications that

[2] See Humberstone (1992) for further philosophical uses of the notion of direction of fit.

[3] Smith, 1994; Humberstone, 1992; Zangwill, 1998; Platts, 1997. For criticisms, see Humberstone, 1992; Copp and Sobel, 2001; Schueler, 1991; Schueler, 2003.

many philosophers have tried to draw from it. Unfortunately, as I argue in section 4, the answer is "no". In particular, I examine whether any notion of direction of fit indeed implies the Humean Theory of Motivation. The Humean Theory of Motivation stands in opposition to the position that I call "motivational cognitivism". Roughly, motivational cognitivism views the relation between (moral) knowledge and moral action as a relation between capacity and exercise; according to the motivation cognitivist, there is some kind of moral knowledge such that its possession guarantees that the agent is motivated to act morally. Is motivational cognitivism compatible with the claim that the theoretical and practical inquiry have distinct formal ends? I argue in section 4 that motivational cognitivism is not only compatible with this claim, but that the idea that theoretical and practical inquiry have different formal ends helps provide an attractive formulation of motivational cognitivism.[4] In particular, I try to show in this section how the notion of direction of fit helps answer a familiar objection to motivational cognitivism. Motivational cognitivism is often presented as the view that accepts the existence of "besires".[5] Besires are supposed to be complex mental states that have the direction of fit of belief towards one content (say *p*) and the direction of fit of desire towards another content (say *q*). However, this very complexity suggests that the motivational cognitivist has simply gerrymandered a state to fit her position. Once we recognize that the mental state is Janus-faced in this way, why couldn't the faces be pried apart? Why aren't we talking about two distinct states that might or might not be co-instantiated by an agent at a time? Our analysis of direction of fit delivers an improved understanding of the kinds of mental states that the motivational cognitivist must postulate and allows her to answer these questions satisfactorily. It allows us to see that the motivational cognitivist is not committed to anything gerrymandered or out of the ordinary; the kind of complexity in question turns out to be no different than the complexity we are committed to accept independently of our views about the nature of moral motivation.

The last section tries to answer an important objection to the argument of the paper. It seems that understood this way, the notion of direction of fit does not fully capture the differences between theoretical and practical reasoning brought to light in Anscombe's example. In particular, it does not account for the fact that there seems to be nothing wrong with

[4] Although it is hard to make claims about past philosophers that are not contentious, I believe Socrates and Kant are clear examples of motivational cognitivists. Among contemporary ethicists, John McDowell seems to be defending such a view in McDowell (1998).

[5] This awkward, but doubtless very useful, label is introduced by Altham (1986).

the propositional attitudes held by the shopper; to the extent that the agent makes a mistake in this case, it, as Anscombe describes it, is a mistake of performance. One might think that it is exactly this feature of the example that should lend support to the Humean Theory of Motivation. As long as this feature is left dangling, one cannot avoid suspecting that even if we have captured *a* notion of direction of fit, we have not captured *the* notion of direction of fit that lends support to the Humean Theory of Motivation. However, I argue that there's something problematic about this characterization of the shopper's attitudes. There *is* something wrong with the propositional attitudes of the shopper; mistakes of performance are best understood as inferential mistakes. What the example reveals, however, is that the content of all-out attitudes in practical reason is always particular, and that, in practical reason, the inference from general to particular is non-trivial in a way that finds no parallel in theoretical reason. And although this is doubtless an important difference between practical and theoretical inquiry, it brings the Humean no support.

2. Working out the Metaphor: The Counterfactual Dependence Strategy

At first, there's something quite intuitive about the distinction between a mind–world and world–mind direction of fit. The basic idea is that a mental state could aim either at tracking the world or at changing it, and to each of these will correspond a different direction of fit; beliefs tend to, or ought to, fit the world, while desires tend to, or ought to, make the world fit them. Explicating these metaphors, however, has been notoriously difficult. One seemingly promising strategy to cash them out is what I call the "counterfactual dependence strategy". The counterfactual dependence strategy tries to explicate the different directions of fit of beliefs and desires by appealing to the fact that beliefs and desires that *p* have different relations of counterfactual dependence to *p* itself or to some third attitude towards *p*. The general idea is to explore the intuition that my belief that *p*, but not my desire for *p*, should be tracking the facts. On the other hand my desire that *p*, but not my belief that *p*, loses its point once *p* has been brought about.

The counterfactual dependence can be either strict or loose. That is, one can claim that the counterfactual dependence always obtains or that it obtains in most or in normal cases. Insofar as one aims to provide an analysis of the notion of belief in terms of direction of fit, appeal to a loose

relation of dependence will probably be of no help.[6] But establishing a loose connection might be enough if one has more modest philosophical ambitions, so I'll leave this possibility open.

The straightforward, but certainly hopeless, version of this strategy would be to claim that belief, but not desire, is counterfactually dependent on *p* itself. Given that we're neither omniscient nor infallible, there could be no such strict relation of counterfactual dependence between the belief that *p* and *p* itself. But a looser relation does not fare much better. Let us take a quick look at a possible suggestion:

(1) Under normal circumstances, *S* would not believe that *p* if it were not the case that *p*

This suggestion faces a dilemma. On the one hand, one can specify "under normal circumstances" so as to make sure that (1) will come out true. If, for instance, we were to understand "normal circumstances" as "circumstances under which a believer is reliable", then (1) would indeed be true.[7] But it's hard to see how to spell out the idea that an agent is reliable in a certain context other than by an appeal to the idea that in such a context the agent tends to believe that *p* only if *p*. More generally, there doesn't seem to be any way to spell out the idea of "normal circumstances" that is non-arbitrary and that makes (1) come out true. Certainly "normal" could not be a statistical notion; there are many common circumstances in which one tends to form false beliefs. We could substitute a normative notion, such as "for appropriately formed beliefs" for the notion of "under normal circumstances", but this would encounter a similar dilemma. On the other hand, we could stipulate that only true beliefs are appropriately formed, but this would trivialize condition (1). Or we could, for instance, identify "appropriately formed beliefs" with beliefs that were formed by rational processes. But here it would seem that one could form false beliefs when one is following otherwise rational processes.[8] Faced with these problems, a wise proponent of the counterfactual strategy should opt for trying to find a counterfactual dependence between belief and a mental attitude that, in some way, is supposed to track the world. However it's unclear that such proposals

[6] Humberstone (1992) places universality as a constraint on the notion of direction of fit. However, whether this is a reasonable constraint depends on the aim of invoking a notion of direction of fit.

[7] Of course, even this claim is a simplification, since a context in which a believer is reliable is not necessarily one in which she'll be infallible.

[8] One can avoid this possibility by identifying rational processes with reliable processes of some sort. But to close off this possibility one would need to understand reliable processes as ones that tend to generate true beliefs, again trivializing condition (1).

can overcome the original dilemma. Take, for instance, Michael Smith's account of direction of fit in these terms:

> A belief that *p* tends to go out of existence in the presence of a perception with the content that *not p*, whereas a desire that *p* tends to endure disposing the subject to bring it about that *p*.[9]

"Perception" can be read in two ways. In the first, "perception that *p*" implies a "belief that *p*". In this sense of "perception that *p*" is not much different than a belief that *p* (or perhaps a specific case of a belief that *p*). In a second sense of "perception that *p*", one can have a perception that *p* even when one does not believe that *p*. Let us start with the first sense. Of course if one's aim is to provide an *analysis* of belief, using a notion of perception that presupposes this very notion is not going to be of much help.[10] But even if one is not intending this to be an analysis of belief, the above proposal cannot be very illuminating if perception is to be understood as being a belief. For all the proposal would say is that belief that *p* is incompatible with belief that *not p*, whereas desire that *p* isn't. This no doubt shows that belief and desire are not the same attitude, and perhaps even that typically desire that *p* and belief that *p* tend not to coexist. Understanding the proposal this way makes it come out true, but not very informative. In fact, the same contrast can be drawn between "belief that *p*" and "supposing that *p*", or "suspecting that *p*", or "wondering whether *p*", but this obviously does not show that these attitudes have a distinct direction of fit, let alone that they have the same direction of fit as desire; it just shows that none of these attitudes can be identified with belief. In fact, even if one is not hoping for an analysis of the notion of belief one would hope at least that the notion of direction of fit would throw light on the different ways in which belief and desire are or should be related *to the world* (or to the facts). But understood in this manner, the notion of direction of fit speaks only to the different "mind–mind" relations that beliefs bear to other beliefs and desires.

I take it that Smith himself intends the second reading of the notion of perception, the one according to which "perception that *p*" does not imply "belief that *p*".[11] But it is not clear how to spell out this notion of perception in such a way as to make Smith's proposal come out true. Copp and Sobel summarize the problem nicely:

> It might seem ... unsurprising that we cannot find an introduced state that counts as in some way a perception with the content that not *p*, that is *not*

[9] Smith (1994: 115). [10] Copp and Sobel (2001) make this point.

[11] In fact, if he did, substituting "belief" for "perception" in the quote above would do just as well, and would be much more perspicuous.

itself a belief, but that interacts with the belief that *p exactly as if* it were an incompatible belief.[12]

After all insofar as a perception that *p* does not imply that one believes that *p*, one could have a perception without forming any tendency to believe. Copp and Sobel give the example of the common optical illusion, in which the asphalt ahead of the driver on a highway might look like a puddle of water. This kind of illusion does not have any tendency to make one believe that there's a puddle on the road; drivers are typically not fooled in any way by it. We can call these optical illusions 'innocent' illusions; although they are cases in which it looks to the subject as if it were the case that *p*, being under this kind of illusion does not make it more likely that the agent will believe *p*. So if the proposal is to come out true, we cannot include this kind of optical illusion as a case of 'perception'. However, it seems hard to find a mental state *M* that satisfies all of the following:

(a) Innocent illusions are not cases of *M*
(b) Being in state *M* with the content *p* does not imply that the subject believes that *p*
(c) The relation between mental states *M* and beliefs with the same content can provide an adequate explanation of the notion of direction of fit.

However, it's not clear that these difficulties are insurmountable. Let us think of a mental state that can be loosely described as "taking *X* as evidence for *p*", or "taking it to be the case that there is reason for *p*", or simply "it appearing that *p*". We can arrive at a somewhat more precise understanding of the state as follows: belief is an "all out" state. That is, believing that *p* is incompatible with believing that *not p*, and there are no states that override one's belief that *p* in the formation of one's unconditional theoretical stance towards *p*. There is no state of, say, "really, I mean it, believing that *p*" for which the belief that *p* provides prima-facie grounds. "Having it appear that *p*" on the other hand is the prima-facie "version" of a belief that *p*. In the absence of countervailing evidence or any reason to think that the appearance is illusory, being in such a state will lead the subject to form the belief that *p*. Using this state in order to understand belief is informative insofar as being in a state such that it appears to the subject that *p* does not imply that the agent believes that *p*. Moreover being in such a state does dispose someone to form the belief that *p*, at least to the following extent: if it appears to a subject that *p* and yet the subject does not believe that *p*, then some explanation is required in terms of countervailing dispositions

[12] Copp and Sobel (2001: 49). See also Schueler (1991) and Humberstone (1992).

or countervailing reasons. Copp and Sobel consider a similar proposal, but they discard it on the grounds that such a state would presuppose a notion of belief. According to them,

> if [the subject] allows that there is evidence for not *p*, she must believe that it counts in favor of believing that *not p*. That is, the judgment that something counts as evidence for *not p* just is a belief.[13]

But there's no reason to think that the state of "appearing that *p*" is such that you are in such a state if and only if you *believe* that there's some *X* such that *X* counts as evidence for *p*. For instance, one can feel oneself in the grip of the gambler's fallacy even if, upon reflection, one believes that there is no reason to think that, say, it is less likely that the next coin toss will be heads, given that the last five ones were heads. In other words, even if I don't believe that the previous tosses count as evidence that the coin will land tails next time, I still *take* the tosses to be evidence for the proposition that the coin will land tails next time. One can say here that the agent holds contradictory beliefs, but it seems more plausible to say that it appears to the agent that it's more likely that the coin will land on the tail side, or that she *takes* the previous tosses to be evidence that the coin is more likely to land on the tail side, even if she does not believe the previous tosses constitute a reason to believe that heads is less likely.

It might be useful to take a look back at the case of the puddle illusion, and see how it would be handled by the proposal we're considering. We can think of two possibilities here. First, it might be the case that the best account of the driver's process of belief formation is something like the following: the driver does take the visual perception to be *some* evidence for believing that there is a puddle up ahead on the road, but the evidence in question is overridden by his knowledge that, under those conditions, such perceptions are likely to be misleading, and the absence of any further reason to think that there are puddles on the road. In this case, the state in question is an appearance of the relevant kind (a prima-facie theoretical attitude), but it is no counter-example to the view that such states always dispose the agent to believe the content of the state. The agent in question does form a disposition to believe that there is a puddle ahead on the road, but it's "neutralized" by countervailing dispositions to refrain from forming the belief in such situations. Or perhaps the best account of the process is one in which the agent never takes the visual perception to be evidence for believing that there is a puddle in the road. In this case, the visual perception does not give rise to a disposition to believe its content. But it is also not a case of being an appearance of the relevant kind; it's not a prima-facie attitude at

[13] Copp and Sobel (2001: 49).

all (even if, typically, visual perceptions are prima-facie theoretical attitudes, this is a case in which one isn't). So, also in this case, the puddle illusion is not a counter-example to the proposal in question. In fact, understood this way, we have a close parallel between, on one hand, appearances and beliefs, and on the other hand, desires and actions. Incompatible desires need to be sorted out on the way to action, and, similarly, incompatible appearances must be sorted out on the way to belief.[14]

However, even though this proposal escapes each horn of the dilemma, it faces a different problem. As I pointed out above, desire is in fact analogous to the state of appearing that *p* rather than to belief. In light of this point, we can say more generally that some attitudes are *prima-facie in character* whereas others are *all-out in character*, or "prima-facie" and "all-out" attitudes for short. Appearances and beliefs are, respectively, prima-facie and all-out attitudes in the theoretical domain. On the other hand, desires and actions (or intentions) are prima-facie and all-out attitudes in the practical realm. We can now say that this revised version of the counterfactual strategy accounts for the notion of direction of fit in terms of two distinct pairs of prima-facie and all-out attitudes that belong to the two distinct realm of inquiries: appearance and belief in the case of theoretical inquiry, and desire and action (or intention) in the case of practical inquiry. However the notion of direction of fit was supposed to characterize exactly what was distinctive to each realm. It does not help to notice that there are two corresponding pairs of attitudes, rather than just one pair, that are candidates for this characterization. In fact, as far as the positive characterization goes, so far we simply assumed that there are two pairs of attitudes rather than just one. Tempting as it is to think that desires and actions (or intentions) on the one hand, and appearances and beliefs on the other hand, must form two distinct relations, the prima-facie and the all-out attitudes, we just *assumed* that they are different; we did not provide any characterization of the difference. It is also tempting to invoke the metaphor of direction of fit here to explain the difference between the two pairs of attitudes, but this is obviously circular.

One can however make progress here by trying to identify distinctive features of the relations between the prima-facie attitudes and the all-out attitudes in the different fields. Theoretical inquiry is the search for what is the case, and practical inquiry is the search for how to act; these different kinds of inquiries might dictate different relations between prima-facie and *all-out* attitudes. In particular, one might want to say that, in theoretical

[14] In Tenenbaum (1999), I discuss this parallel in more detail and argue that even the ill-formed belief in the case of the gambler's fallacy finds a parallel in the case of practical reason; I argue there that we should understand *akrasia* in similar terms.

inquiry, prima-facie attitudes are (should be) taken up in all-out attitudes insofar as the agent accepts (should accept) that the content represents how things are, whereas in practical inquiry, prima-facie attitudes are (should be) taken up in all-out attitudes insofar as the agent accepts (should accept) that the content represents what he or she is to do. To say that one's inferences from prima-facie attitudes in theoretical reason to beliefs is or should be guided by *how things are* is to say that the process of belief formation is in some sense guided by the ideal that one's beliefs should be guided by the truth. In other words, this proposal now postulates some kind of constitutive relation between belief and truth; roughly speaking, a belief is an attitude whose formation is, or ought to be, guided by the pursuit of truth. We can say in this case that truth is the *formal end* of inquiry. It's an end that guides, or ought to guide, every instance of engaging in theoretical inquiry.[15] And one might propose that desire and intention, on the other hand, bear a similar relation not to the truth, but to something else. This something else could be 'the good', 'the desirable', 'rational action', 'autonomous action', or something else. I'll assume that what ought to guide us in those transitions in practical reason is 'the good',[16] but again, the argument does not hang on this being the correct choice.

This is indeed a promising proposal. But one must note that this proposal leaves the counterfactual dependence strategy behind; we end up trying to capture the notion of direction of fit in terms of the constitutive relation between belief and truth. The counterfactual dependency strategy collapses into a strategy that tries to capture the direction of fit in terms of, on the one hand, the relation between belief and truth, and, on the other hand (if I am correct about what the formal end of practical reason is), the relation between desire and the good. We can now simply ask what (if any) implications flow from the fact that belief and desire have different formal ends. But before we can answer this question we need to clarify what it means to say that belief is, or ought to be, guided by truth, whereas desire and intention are, or ought to be, guided by the good.

[15] Of course I cannot do full justice to the various issues surrounding the notion of a formal end of inquiry here. I discuss these issues in more detail in Tenenbaum (forthcoming (*a*)). However, a few words of warning might be important. I am not using the notion in the same way as Velleman does (2000*c*), at least insofar as Velleman thinks that specifying the formal end of an inquiry is completely uninformative. If anything, the notion is closer to what he calls there the "constitutive aim" of inquiry or belief. However, it is not quite the same notion either since I am not committing myself to the view that the formal end of inquiry could be *fully* understood apart from its being what constitutes successful inquiry. For an illuminating discussion of these issues, see Clark (2001). The formal end of an inquiry is in my view what Clark calls a "generic object".

[16] Mostly because I think that this is the correct view. See Tenenbaum (forthcoming (*a*)).

3. Aiming at the True and the Good

Although most philosophers agree that belief bears a certain constitutive relation to the truth, the characterization of this relationship is no easy matter. To say that the relation is constitutive is to say that nothing can count as a belief that *p* unless it stands in this relation to the fact that *p*. And, of course, as long as we understand the relation in this manner, we must use some expression like "the aim of belief" in spelling out the condition; a belief that *p* can easily coexist with the fact that *not p*. One can say that believing *p* implies holding *p* to be true. But the truth of this statement depends on what we mean by "holding true". There is a sense in which we hold a statement to be true when we assume something for the sake of argument,[17] and one can certainly believe things without explicitly considering the matter in terms of the truth of a sentence or a proposition. I'm not sure there is any non-trivial way to characterize the relevant sense of "holding true". We can say things like "holding with endorsement" or "holding with acceptance", but if "endorsement" and "acceptance" don't just mean "belief" in this context, it'll be easy to build counter-examples to the claim that to believe is to hold with endorsement or acceptance.[18] For our purposes, it suffices to say that believing *p* implies holding *p* to be true in the sense of "holding true" characteristic of belief. In any case, put this way, this is not quite a characterization of truth as the aim of belief. After all, to say that I hold something true is not to say that I hold it true because I aim to hold it true;[19] it certainly does not follow from the fact that one holds x to be y that one aims to do so. The idea that belief aims at the truth is more robust than the idea that believing *p* amounts to, or implies, holding *p* to be true in a certain way. But how should we understand what the postulation of such a constitutive aim adds to the idea that believing implies holding true?

One way to understand this addition is to think that an agent does not count as believing that *p* unless the agent forms the belief guided by the aim of believing truly. Now this idea needs some refinement. Obviously we do not form beliefs by engaging in explicit instrumental reasoning about maximizing our chances of hitting the truth for every single belief we have. We need to understand having this end in a way that does not

[17] Velleman (2000*b*).

[18] Cf. Van Fraassen's (1980) distinction between "accepting" a theory and "believing" a theory.

[19] This is indeed Velleman's characterization of belief (2000*b*). It'll become clear momentarily why I think that this is not an adequate characterization of belief.

imply anything so obviously false. Various things are often said with respect to belief that could be helpful in understanding better what it is to have such an end. One can say, for instance, that we cannot form beliefs at will,[20] and thus that we cannot have any aim other than believing truly in forming beliefs. Or, one can appeal to Moore's paradox[21] to explain the impossibility of forming a belief in disregard of what one considers to be true. But probably the clearest and most promising way to spell this idea out is to say that belief must respond to evidence;[22] that is, no state counts as belief for *p* if the state is not responsive to the evidence for or against *p*. Again, the notion of "responding to evidence" needs to be clarified here. It's obviously false that all our beliefs are proportioned to evidence, and some of our beliefs, especially unconscious beliefs, beliefs that are the result of wishful thinking, etc., are formed by processes that are in no way truth-conducive.[23] The condition of responsiveness to evidence should be something like the following:

(RE) An agent counts as believing that *p* only if the agent does not consciously hold the belief due to non-epistemic reasons.

This is a relatively weak requirement. (RE) allows that some beliefs be held for no reason. It also exempts unconscious beliefs, and beliefs that result from self-deception and the like, from its "evidentialist" requirements.[24] Ideally, we would spell out what is meant by "epistemic reasons" and what counts as a belief being due to a reason rather than another. But since spelling out would cost some generality, I will leave this task to the side. The idea that something like (RE) must be true is initially very plausible, but I think it cannot withstand scrutiny. (RE) is particularly plausible if it is understood as part of a *general* condition on an agent *having beliefs*; it is plausible to assume that no one can count as a believer if his beliefs do not satisfy the consequent of (RE) often enough. However, as a specific condition of what makes a *particular* mental state a case of an agent having *a* belief, I think it is false. Here are a few counter-examples to the *specific* condition:

1. Mary is up for the job of her dreams. She looks at the ad, and she's struck by the thought that she is a shoo-in for the position. But Mary thinks that the experience of failing to get the job of one's dreams after expecting that one would is so painful that it's better not to

[20] Williams (1973). [21] Railton (1997).
[22] For instance, Wedgwood (2002). [23] See Shah (2003).
[24] To make matters simpler, I'm also leaving aside the fact that someone defending (RE) would also want to include a further requirement to the effect that an agent does not believe *p* when she is in possession of overwhelming evidence for *not p*.

believe that she'll get the job. She decides to persuade herself that she'll not get the job.

2. Clara's husband has been indicted for a crime. All the evidence points to his guilt. However, Clara doesn't believe that her husband is guilty. When confronted with evidence, she says: "I trust my husband, and to trust somebody involves being committed to believe his innocence, even when the evidence warrants the opposite conclusion."
3. Otto's son is missing. All evidence points to the child's death. Otto acknowledges this fact, but he says: "I can't just let go of him like this. I must continue to believe that he's alive (and thus I believe that he's still alive)."

Now these are cases in which, although the agent in question still seems to hold the propositions in question to be true, they are not cases in which the agent tries to form the belief that *p* only if the evidence warrants the belief. They are also not Pascal-like cases in which the agent forms a plan now to ensure that in the future she'll find warrant for a certain proposition ((1) is the closest to this case).[25] These are beliefs that are not *currently* being sustained by any kind of aim of maximizing the chances that the belief is true or in accordance with the evidence.

These are all cases in which the formation of belief is guided by goals other than truth. To the extent that (1)–(3) are compelling, the idea that belief must be guided by truth is descriptively false.[26] Of course these examples do not conclusively establish this point. One can, for instance, try to explain away these cases as cases in which the agent behaves *as if she believed*, but not cases in which she really believes the statement in question. Or one could say that these are cases in which the agent in question uses some kind of non-standard evidence. The father takes a certain gut feeling as evidence, or the spouse her special acquaintance with her partner as evidence. I don't find these replies promising; I don't think that one can escape the conclusion that one does not always aim at the truth when forming beliefs. Although I can't argue in detail for these conclusions, I hope these examples suffice to give us some reason to think that it would be best to account for the distinctive direction of fit in terms of a *normative*

[25] It is worth noting that faith-based belief in God often seems explicitly to run afoul of (RE). Yet it would be hard to say that people who claim to believe in God while acknowledging that there's no evidence for their existence don't have a proper belief. I owe this example to Fred Schueler.

[26] This needs some qualification. It might be correct to say that I can't count as believing *p* if my belief is supposed to survive overwhelming evidence to the contrary. I'm not going to try to settle this issue.

relation between belief and truth. In particular, we can say that believers are under something like the following normative requirement:

(NR) Believe *p* only if *p* is true.

The kind of normative requirement in question needs to be stated with caution. Cases (1)–(3) do not clearly involve an agent who is doing anything that is, all things considered, wrong or irrational. It might be that trust does require that we override evidence, or that one is better off expecting the worse, or that only a heartless parent would accept anything that far from Cartesian certainty to form the belief that his child has passed away. The normative ideal in question must be an ideal for belief considered solely from a theoretical perspective, insofar as we are engaged in the search for truth abstracted from any other concerns. (NR) is thus best understood as a normative claim about how belief ought to be responsive to evidence *insofar as an agent is engaged in theoretical inquiry*. Of course, the closer one is to accepting the view that belief about *p* should be understood simply as an agent's all-out attitude insofar as she is engaged in theoretical inquiry, the more stringent one's interpretation of "often enough" will be, in our claim above that in order to count as a believer one needs to satisfy the consequent of (RE) often enough. But (1)–(3) should make us suspect that "often enough" cannot become "always".

Obviously there is room for refining (NR), but the simple version of the requirement should suffice for our purposes.[27] To say that (NR) should guide one's belief formation at least insofar as one is engaged in theoretical inquiry is to say that in theoretical inquiry moves from prima-facie and all-out attitudes, as well as, obviously, moves from all-out to all-out attitudes in theoretical reasoning are guided by the ideal of truth, and, roughly, inferences are judged appropriate to the extent that they are truth-conducive.[28] To the extent that practical inquiry has a formal end, a similar thing can be said about it. Moves from prima-facie to all-out attitudes in practical reason are guided by the ideal of the pursuit of the good (assuming, again, that this is the formal end of practical reason), and inferences are judged appropriate to the extent that they are, in some sense, "good-conducive".

We can now refine the understanding of direction of fit proposed in the previous section: as we move from, say, a certain perception to a belief, insofar as what we are engaged in can count as theoretical inquiry, we

[27] See Wedgwood (2002) for a normative account of the relation between belief and truth.

[28] Or likely to be truth-preserving. Of course, one needs to make room for inferences that are not necessarily truth-preserving.

should be guided by the truth-conduciveness of the move. This will count as unsuccessful theoretical inquiry if the belief formed is not true or if the inference was not truth-conducive. On the other hand, when we form intentions on the basis of our desires, we should be guided by the "good-conduciveness" of the move; that is, by the fact that acting (or intending) on the basis of such desires counts as performing actions that are good (or, in other words, as acting well). This will count as unsuccessful practical inquiry if the action was not good (it would have been better to have acted differently), or if the inference was not good-conducive or warranted (the move from one's desires to the action was unwarranted). Ultimately the claim that belief and desire have different directions of fit is best understood as the claim that inferential moves in practical and theoretical inquiry are guided by distinctive formal ends.

One might object that while my description of the relation between belief and truth borders on triviality, the similar relation between intention or action and the good does not obtain. Many believe that one often acts without in any way pursuing what is good, and I have given no argument against their position. This is an important issue and I cannot do full justice to it here.[29] But I hope that the following remarks will show that this objection is not as worrisome as it might appear. First, just as we allowed that in the case of theoretical reason there might be belief formation that is not actually guided by truth, we could also allow that some actions are not guided by the good; perhaps, this is how one ought to understand *akrasia, accidie*, etc. However these actions would be, on the view proposed, in some way defective, by failing to conform to the formal end of practical reason, in the same way that a belief whose formation is not guided by the pursuit of the truth is defective as a piece of theoretical inquiry.[30] One might object here that one can act in a way that is unimpeachable and yet not in the pursuit of anything that one considers to be good. In fact, one might most fully identify with "perverse" pursuits, and feel "alienated" when one is pursuing something that one takes to be valuable.[31] Again, discussing this topic in any detail would lead us far astray. Obviously, if we accept that the good is the formal end of practical reason, we will doubt the coherence of this way of describing any piece of human behaviour. So these claims are best constructed as proposing that the good is not the formal end of

[29] For an extensive defense of the view that the good is the formal end of practical inquiry, see Tenenbaum (forthcoming (*a*)).

[30] Notice that although, as I pointed out above, it might be legitimate to form a belief on pragmatic grounds, it is hard to see how we can have a coherent view of an intention formed by any grounds that do not pertain to practical reason or practical inquiry.

[31] See Velleman (2000*b*, 2000*d*).

practical inquiry; indeed, philosophers who think that this kind of action is a real possibility typically think that something else is the formal end of agency.[32] But, as I said above, the argument of the paper does not depend on taking "the good" rather than "autonomy" or something else to be the formal end of practical reason.

4. Direction of Fit and the Humean Theory of Motivation

Let us distinguish two versions of cognitivism in ethics. Some cognitivist views might accept that the virtuous person's motivational state is *causally* related to a cognitive state, such as an evaluative belief, and deny that being properly motivated is a further cognitive achievement. The motivation itself is a blind disposition that may fail to be effected by the existence of the relevant evaluative belief, even if it is typically effected, or it is effected insofar as the agent is rational.[33] A more stringent form of cognitivism, however, would hold the view that the virtuous person is in a cognitively superior state than the vicious or the akratic person, and that the motivational state of the virtuous person is itself a cognitive state. I will call the latter view "motivational cognitivism"; according to motivational cognitivism, moral motivation stands to our rational powers as exercise to faculty. For motivational cognitivism, if an agent doesn't act as the virtuous agent would, then she cannot be credited with the same understanding of morality that the virtuous agent has. She might fail to have the same beliefs that the moral agents have, or her grasp of the content of the beliefs might be defective, or perhaps she does not fully understand the grounds for forming the relevant moral beliefs.[34] For the motivational cognitivist, differences in motivational states must be fully accounted by differences in cognitive states.

[32] I take it that this is a correct description of Velleman's position in the matter. For Velleman the constitutive end of action (and I take it, a fortiori of practical inquiry) is autonomy, not the good. See Velleman (2000*c*, 2000*d*).

[33] See Smith (1994). Even though Smith claims that an agent, insofar as she's rational, will have her motivation lined up with her values, what makes her a rational agent on Smith's view, as far as I can see, is simply the fact that this causal relation obtains. Smith insists that the agent who suffers from *accidie* or *akrasia* does not necessarily lack any kind of knowledge available to the virtuous agent. Rob Shaver (n.d.) argues that Sidgwick also held a view of this kind.

[34] The possibility that the difference between the virtuous agent and the non-virtuous agent lies in the grounds of their beliefs, instead of the beliefs themselves, is often strangely absent in discussions of the topic. Of course if only beliefs can ground beliefs, then one difference reduces to the other, but if I am right, an adequate analysis of direction of fit presupposes that this claim is false.

I'll assume that if the direction of fit metaphor lends support to the Humean theory of motivation, it must be at least be capable of showing that motivational cognitivism is untenable. One can at least say that if the Humean theory of motivation accommodates motivational cognitivism, there's nothing left distinctively "Humean" about it; there's no concession left to be made to those who stand on the other side of the fence, especially because anti-Humeans are often happy to grant that one can always ascribe some kind of desire to every case in which an agent acts for whatever reason.[35] Motivational cognitivists are often described as those who think that belief alone, and moral belief in particular, can motivate. This is certainly one way in which one could endorse motivational cognitivism: a belief about reasons for action in a certain situation, or an evaluative belief would by itself generate action. We can call this view "belief-based (BB) motivational cognitivism".

However, in focusing on the debate about the truth or falsity of moral judgments one overlooks another possible form of motivational cognitivism suggested by our discussion. Let us grant that, on the side of theoretical reason, the relation between prima-facie and all-out attitudes ought to be guided by the ideal of believing the true—according to inferential patterns dictated by this ideal—and that successful cognition will require a certain kind of non-accidental relation between those attitudes and what is actually true. One could adopt a parallel view about the nature of practical reason. One could say that the relation between prima-facie and all-out attitudes ought to be guided by the ideal of pursuing the good—according to inferential patterns dictated by this ideal—and that successful cognition requires a certain kind of non-accidental relation between those attitudes and what is actually good. If one adopts this latter view and if one thinks that moral action (necessarily) bears the right relation to the good, a relation parallel to the relation between knowledge and truth, one accepts a form of motivational cognitivism that is not committed to the view that beliefs can motivate by themselves. I take it that, for instance, Kant held this kind of motivational cognitivism. Kant maintained a sharp distinction between practical and theoretical reason,[36] taking them to be guided by different and irreducible ideals. Imperatives and maxims are our guides in acting. They're certainly not beliefs, and yet they can be cases of successful (or failed) cognition. As if the previous label weren't enough of a monstrosity, I'll call this view "non-belief-based motivational cognitivism" (NBB). It should be obvious that the relation between belief

[35] See McDowell (1998) on "consequential" desires, Platts (1997) on "trivial" desires, Schueler (1995) on "pro-attitudes", and on "motivated" desires (Nagel, 1970).

[36] See, for instance, Kant (1998: B830–1).

and truth, as I have presented it, poses no threat to NBB motivational cognitivism. After all, this form of cognitivism *advocates*, in this regard, a sharp separation between theoretical and practical reason. But can BB motivational cognitivists accept this analysis of direction of fit? After all, most arguments for the Humean Theory of Motivation target the idea that *belief* could motivate.

If BB motivational cognitivists are right, then some of our beliefs are not only theoretical attitudes but also practical attitudes. As such they would have to have both directions of fit at once. Arguments for the Humean theory of motivation based on the notion of direction of fit try to show that it is either incoherent or very implausible to think that the same attitude could have both directions of fit.[37] Now the Humean might start his argument by claiming that it is incoherent to have an attitude with both directions of fit towards the same content *p*. This claim does not directly contradict any kind of motivational cognitivism. The motivational cognitivist thinks that some beliefs with contents such as "it would be good to help the little child" or perhaps "the little child needs help (and nothing prevents me from helping her)" are inseparable from a motivation to *help the child*. No motivational cognitivist thinks that the moral agent rather has the absurd motivation to bring about the very content of these beliefs. But once we present the motivational cognitivist this way, we seem to provide the advocate of the Humean Theory of Motivation with a powerful argument against BB motivational cognitivism. The mental state that the BB motivational cognitivism postulates turns out to be a rather complex state; what Altham calls a "besire". These mental states are composed of two different contents and two different attitudes, corresponding to each direction of fit, for each of these contents. The agent is supposed, at the same time, to believe the content "it is good to help the child" and be disposed to bring about the content "I help the child". But if this is so, what could be the grounds for claiming that they are inseparable? Why couldn't an agent have one half of the besire but not the other? Moreover isn't this exactly what happens to certain agents, especially agents suffering from motivational disorders such as dejection, *accidie*, or depression? Don't they, say, continue to believe that it would be good to help the child, but fail to garner motivation, or at least sufficient motivation, to bring about that they help the child?

Our analysis of the notion of direction of fit should help us understand why we should not be persuaded by this argument against BB motivational

[37] I have in mind here in particular, Smith's arguments (1987; 1994: ch. 4). However, I am presenting the arguments in a slightly modified form.

cognitivism. It is worth first noting that our discussion suggests that the term "besire" is ambiguous; one could be advocating a view about the existence of any of the following:

(i) a mental state that is both a theoretical and practical all-out attitude;
(ii) a mental state that is both a theoretical and practical prima-facie attitude;
(iii) a mental state that is both an all-out theoretical attitude and a prima-facie practical attitude;
(iv) a mental state that is both an all-out practical attitude and a prima-facie theoretical attitude.

The above argument for the Humean Theory of Motivation is probably at its best when challenging the existence of attitudes described in (i). But if one wants to argue for the impossibility of any attitude that has multiple directions of fit, one has to show that all attitudes described in (i)–(iv) are incoherent; there can be no such "necessary union of direction of fit". However, as we look into all these possibilities, the prospects for making a case for the incoherence of any case of multiple directions of fit become quite dim. Let us look at an example of an attitude that seems to fall squarely into (iv). Intentions seem to be good candidates for being all-out practical attitudes.[38] Now it seems that forming an intention to ϕ serves as grounds for one's belief that one will ϕ,[39] and so here we have a case of (iv). Now various views on the nature of the relation between belief and intention might make it easier to accommodate the view that there are two separable mental states corresponding to the two directions of fit.[40] However it is hard to believe that general considerations about the formal ends of practical and theoretical inquiry should settle among our views about the nature of intentions. Similarly, let us look at the state of being in intense pain. Arguably, being in pain is not a representational state, and thus it does not have any direction of fit. However, it would be a respectable philosophical position to think it is constitutive of this state, at least in the case of human beings, that the following obtain: (*a*) the agent is at least inclined to believe that he is in pain; (*b*) the agent has some motivation

[38] I actually think that only intentions in action are all-out practical attitudes, but this does not affect the argument. See Tenenbaum (forthcoming (*b*)).

[39] Davidson (1980) famously argues that one can intend to ϕ without believing that one will ϕ. But my claim is much weaker. I only claim that intending to ϕ is inseparable from a prima-facie attitude to ϕ.

[40] Some people think that intentions simply are beliefs. See, for instance, Harman (1976). On this view, this is not going to be a case of (iv). But then one's denial of the existence of besires is hostage to a controversial view about the nature of intention. For various problems that the view that intention is belief faces, see Bratman (MS).

not to be in this state.[41] On this view, being in pain is a mental state of kind (ii). Now one can argue against (*a*) or (*b*) being constitutive of intense pain. But it would be bizarre to try to argue that *not both* (*a*) or (*b*) could be constitutive of pain solely on the grounds that theoretical and practical reason have different formal ends. But why would the situation be different with the motivational cognitivist? Why would it be possible to rule out in advance the possibility that an evaluative belief can also be a practical attitude? The postulation of mental states with multiple directions of fit is not an *ad hoc* maneuver on the part of motivational cognitivism; it is something that we might already be committed to in completely different contexts.

It is worth noting that even the most stringent form of motivational cognitivism needs to be committed only to the existence of states of kind (iii). And our analysis should make it clear that accepting states of kind (iii) does not amount to accepting an attitude that is somehow unique or extraordinary. It is easier to make both points if we start from the obvious fact that a belief can serve as evidence for another belief. Take, for instance, Anita's belief that the indentations in the sand that she's observing right now are tiger footprints. It thus appears to Anita that tigers have been around. Now one might suggest that in this case we have an all-out theoretical attitude that is also a prima-facie theoretical attitude of a different content. The belief "the indentations in the sand are tiger footprints" and the "appearance" with the content "tigers have been around", one might argue, are one and the same state. Opposing this suggestion, one might argue that we should keep the two states apart; one might want to insist that it is at least conceptually possible that one forms the belief without having any attitude, prima-facie or all-out, with the content "tigers have been around". I must confess I find it hard to wrap my mind around the idea that this is indeed a conceptual possibility.[42] I don't see how one can have a full grasp of the content "the indentations in the sand are tiger footprints", have a belief with this content, and yet not have it at least *appear* to him that tigers have been around, given the close conceptual connection between "x is a tiger footprint" and "x is the effect of a tiger's paw making contact with the surface". But if one wants to insist that the separation is

[41] Christine Korsgaard's view on the nature of pain, although different from the view described here, does seem to incorporate motivational and cognitive elements as constitutive of pain itself. See Korsgaard (1996: lecture 4).

[42] I am ignoring an irksome complication. One could produce a footprint in the absence of tigers; one could press a severed tiger paw against the sand. Perhaps someone who sees this footprint knowing how it is produced doesn't take this to be any kind of (overridden) evidence that tigers have been around. However, one could complicate the belief so as to rule out this possibility.

conceptually possible, one can just replace this example with one of a closer conceptual connection between the content of the belief and the content of the appearance. Perhaps "John has a sunburn" for the belief, and "John was exposed to the sun" for the appearance. At any rate, it'll be hard to argue against the following general claim of conceptual connection:

(CC) For some distinct contents X and Y, if a subject S fully grasps X and Y, then it is necessarily the case that if S believes X then it appears to S that Y (S has a prima-facie theoretical attitude with content Y).

I know of no general reason to think that one can rule out that at least some belief states stand in this kind of relation to other beliefs states for which they are evidence. It is also important to note that nothing I said above rules out the possibility that the appearance is conclusive. By a "conclusive appearance", I mean something along the lines of "providing obviously conclusive evidence"; if one has something that counts as obviously[43] conclusive evidence for *p*, and one understands the evidence, and that it is conclusive evidence for *p*, arguably, one necessarily forms the belief that *p*. Similarly, if someone has a prima-facie attitude of content *p* that is (obviously) conclusive, one will necessary form the belief that *p*. One might argue that the belief that John has a sunburn doesn't imply only that it appears that John has been exposed to the sun, but, in fact, the appearance in question leaves no room to doubt that John has been exposed to the sun; in this case, once one believes that John has a sunburn one cannot stop short of the belief that John has been exposed to the sun. Again, here one might think that this is not true for this example, and one might doubt whether it is true for any example. All that I want to note at this point is that one cannot rule out in advance the possibility that having a belief state that X will imply having a belief state in which it appears conclusively that Y.

Now one might say that (CC) does not imply that the belief and the appearance are one and the same mental state; for some reason, one might want to say that they are two states such that one could not be in the former without being in the latter. This might be a plausible move, and since for our purposes this does not make much difference whether the move is made or not, I'll just talk about one state conceptually implying the other, without prejudging whether we have one or two mental states.

[43] This qualifier should make the demand much weaker than a demand for closure. Smith (forthcoming) suggests that the motivational cognitivist is committed to accepting deductive closure. But I hope it'll be clear that motivational cognitivism is not committed to anything that strong.

If one grants (CC), one grants that an all-out theoretical attitude can entail a prima-facie theoretical attitude. As we saw above, we know that it is possible that some practical attitudes imply the existence of some theoretical attitudes. What reason can we have now to deny that a certain all-out theoretical attitude could entail certain prima-facie practical attitudes? Why couldn't the content of the all-out theoretical attitudes of a virtuous person be such as to imply a certain prima-facie practical attitude? That is, why couldn't the relation between the beliefs of the virtuous person and the desire to act in certain way be just the same as the relation between S's belief that John has a sunburn and the fact that it appears to S that John was exposed to the sun? After all, the BB motivational cognitivist need be committed to nothing more than the claim that having the kinds of beliefs that the virtuous agent has will necessarily motivate. A relatively weak version of motivational cognitivism need not say that moral beliefs necessarily lead to action. But in fact, there is no reason to think, purely on the grounds of the nature of these attitudes, that theoretical all-out attitudes could not entail conclusive appearances in the practical realm. If one cannot rule out the existence of these relations within the theoretical realm, why should it be impossible that a similar relation obtain across realms? An example might help make out this point. Suppose one thinks that "John has a sunburn" conceptually entails a conclusive appearance to the effect that John was exposed to the sun. It is now the case that I cannot attribute to Larry a belief, or at least a non-defective belief, with the content "John has a sunburn", unless I am prepared to attribute to him also the belief with the content "John was exposed to the sun". But if this is so in the case of the relation among beliefs, what reasons do we have to rule out the possibility that certain beliefs can be attributed to the agent only if he is prepared to act in certain ways (or form certain intentions)? This still falls short of a commitment to (i), since the moral belief would probably not suffice to give rise to a full-blown all-out attitude; it would probably lack content to specify in detail the actual intention with which the agent acted. However even the most radical motivational cognitivist need not be committed to anything stronger.[44]

One could insist that it is simply implausible to suppose that certain beliefs are capable of inclining the agent to pursue anything. But this is not an *argument* for the Humean Theory of Motivation; it *is* the Humean Theory of Motivation. More plausibly, one can think that states such as *accidie* or depression speak against the fact that moral beliefs can be conceptually connected to the relevant practical attitudes. After all, the

[44] More on this issue in the next section.

agent who suffers from these ills might have the exact same belief as the virtuous agent. The depressed agent, just like the virtuous agent, could believe that it would be very good indeed to help the poor, but just fail to garner the motivation to do it. It would be *ad hoc*, the Humean may say, to deny that the depressed agent has the same belief just because he fails to act in the same way.

But is it *ad hoc*? Our above discussion should suggest that the answer is "no". The defining thesis of BB motivational cognitivism is the claim that the very fact that motivation is not present is what *makes* it the case that we cannot attribute the full-blown moral belief to the agent in question; motivational cognitivism is not committed to the claim that for every case that the motivation is absent we will have an independent reason not to ascribe the full-blown moral belief to the agent. Of course if the central argument of moral cognitivism were the claim that all those who honestly assent to moral claims behave morally, cases of *accidie* and depression would present a serious challenge to the view. But no motivational cognitivist would defend her view in this manner. Motivational cognitivism takes as its starting point the attractiveness of a picture of morality in which moral activity *is* a form of knowledge.[45] So the motivational cognitivist is committed to seeing those motivational failures as in themselves failures to fully grasp the content of one's moral beliefs, or somehow failing to have the same kind of moral beliefs as the moral agent. Of course, one can dispute these claims. But just as in the case of intention or of being in intense pain, what settles the debate is who provides us with the best conception of the virtuous agent, not considerations about the formal ends of theoretical and practical reason.

One should point out that the BB motivational cognitivist is often saddled with a "molecularist" picture that makes her view seems particularly implausible. The BB motivational cognitivist does not need to claim that the difference between the virtuous agent and the one suffering from *accidie* must be present in *each* belief, considered on its own, that fails to motivate the dejected agent.[46] The motivational cognitivist is not committed to the claim that, when the agent suffering from *accidie* says "I should not be just lying in bed", there is really some part of "not" that he doesn't understand. The BB motivational cognitivist thinks that full understanding of the moral

[45] This is true both of historical figures and contemporary philosophers. Kant says that wisdom (*Weisheit*) is primarily a matter of acting. See Kant (1998*b*: G 405). Among contemporary philosophers, John McDowell (1998) explicitly presents the claim that virtue is knowledge as a motivation for his view.

[46] This molecularist interpretation of the BB motivational cognitivist is certainly encouraged by characterizing the position as one that accepts the existence of besires.

facts ensures actions, but the lack of understanding need not be attributed to a belief considered in isolation. In considering the differences between fully virtuous agents and all sorts of other agents, within the confines of BB motivational cognitivism, we can appeal not only to differences in the contents of their beliefs, but also to differences in the relevant prima-facie attitudes, in how they jointly ground the belief in question, in how they cancel other prima-facie attitudes that seem to undermine the belief, etc. We can see now that the BB motivational cognitivist is committed to something weaker than what we've been suggesting; all she needs to accept is that there are some "packages" of all-out attitudes grounded on certain prima-facie attitudes, such that full understanding of how the whole package hangs together is conceptually connected to a prima-facie practical attitude. It is not implausible to think that there is *something* in this package that distinguishes the virtuous agent from, say, the dejected agent. Considerations of the different directions of fit of belief and desire certainly can give us no reason to be suspicious of this commitment; in fact the considerations show that similar relations hold in other domains.

5. The Detective and the Shopper

Anscombe's example seems to suggest a sharp, independently conceived distinction between the two directions of fit, a distinction that does not seem to be captured by the idea that practical and theoretical reason might have different formal ends. And one might suspect that we failed to find an argument for the Humean Theory of Motivation simply because we failed to capture something important in Anscombe's example. In this section, I'll try to lay this suspicion to rest. Let us go back to Anscombe's example. Let us call the detective Jenny and the shopper Leo, and suppose that that's how things look like:

Leo's List	*Shopping Cart*	*Jenny's List*
Grapes	Cherries	Cherries
Apples	Apples	Plums

We can describe what goes wrong with Leo as follows:

(1) Leo wants to buy grapes.
(2) Leo buys cherries.

On the other hand, we can describe what goes wrong with Jenny as follows:

(1) Jenny believes there are plums in Leo's shopping cart.
(2) There are no plums in Leo's shopping cart (there are apples).

We can notice a few things now. First, Leo's mistake can be characterized as an inferential one[47] in a broad sense of 'inferential'. The mistake was moving from an attitude (a desire) to another attitude (acting with an intention) for which the first was supposed to provide grounds. Also, by saying that this was an inferential mistake, I'm not claiming that Leo was irrational, or that his inferential patterns are blameworthy. All that I am claiming is that he moved from an attitude that was unimpeachable to one that was not. If the pattern of inference cannot guarantee that one always move from unimpeachable attitudes to other unimpeachable attitudes, the agent might arrive at mistakes while being perfectly rational. Acting with an intention is not typically characterized as an attitude. However, whether we can count it as an attitude or not is not particularly important for my purposes, as long as one grants that one does things on the grounds of certain desires or intentions; what I've been characterizing as an inferential relation is just the grounding relation between the desire or intention and the action. We can present the inferential relation in our example as follows:

(3) Leo wants to buy grapes.
(4) Leo acts in such a way as to bring it about that these fruits are in the shopping cart.

Note that, if this is the correct characterization of the inferential relation, Leo's mistake is one that can be located in his moving from a prima-facie attitude to an all-out one, and from a desire with a general content to a particular action.

But note that we can also characterize the detective's mistake as an inferential one in this broad sense of 'inferential',[48] in the move from (5) to (6) below:

(5) It appears (perceptually) to Jenny that these fruits are in the shopping cart.[49]
(6) Jenny believes that there are plums in the shopping cart.

Once we think about the differences in these terms, Anscombe's case is also a case in which the difference between the practical and theoretical cases is a difference between the different formal ends that guide the moves from

[47] Someone might protest that this is a mistake in *performance* not an inferential one. I come back to this point in a moment.

[48] This characterization does not rule out the possibility that the belief is 'non-inferential' in a narrower sense of 'inferential'; that is, it is not inferred from other *beliefs*.

[49] I don't mean to imply that content of perceptual experience must be conceptual. It doesn't matter for my purposes if the inference starts from conceptual or non-conceptual content.

prima-facie to all-out attitudes. However it is hard to shake the feeling that there is something different here, that we have not captured the idea that Leo's mistake must be located in how the agent changes the world, rather than in how the agent changes his mind. It is easy to try to dismiss the difference as just the result of the fact that practical reason concerns action; hardly something that any philosopher has missed. Yet we must acknowledge that there is an important disanalogy between theoretical reason and practical reason that comes up in this example that our analysis of direction fails to capture. The move from the general to the particular in theoretical reason in forming a judgment tends to be trivial. Although subsuming a particular under a concept and forming judgments of the form *Fa* is in no way trivial, moving from judgments of the form *(x)Fx* to judgments of the form *Fa* certainly is. That is, leaving aside very complexly formed predicates and other complications, if my judgment that *(x)Fx* is correct, there won't be much room for mistake in moving from the general judgment to the particular one. However the same is not true in the realm of practical reason. Of course, there's no agreed-upon equivalent of universal instantiation in the realm of practical reason. But without trying to work out the details of this proposal, we can think of the move from the general intention to a particular action as a similar move. One example of this kind of inference would be the following: I infer from my thought that, all things considered, actions in which I pause for a moment and draw a circle in the air are desirable to my acting so as to bring it about that I'm drawing a circle in the air right now in a particular way. However, unlike the case of theoretical reason, this move is in no way trivial, for despite the simplicity of the predicate "drawing of a circle", there's no guarantee that I'll succeed in actually drawing a circle in the air; in fact, I'll probably fail. Practical reason allows for mistakes of *performance*,[50] mistakes in trying to execute a flawlessly formed, simple intention.

Arguably all-out attitudes of practical reasons are always particular judgments; they are cases of acting with an intention as described in statements such as (4). Given that an intention does not by itself determine how it will be carried out, an intention that is not an intention in action will always leave room for revision as one tries to carry out the intention in concrete actions. Insofar as practical reasoning aims to issue in some kind of action, forming a general intention is still being in a state that falls short of being an all-out attitude. Any such general intention must have some *ceteris paribus* conditions that could fail to obtain, and thus fail to be the agent's final view about how she shall or should act. Therefore such states are not

[50] This is how Anscombe herself (1963) identifies the mistake of the shopper.

all-out attitudes. If one does not want to go all the way to the Aristotelian view that the conclusion of practical reason is an action, one will need at least to say that the conclusion is a decision to engage in *this particular action.*[51] If this is correct, mistakes of performance are failures that can be coherently ascribed to *any* all-out practical attitude, but to no theoretical attitude. In sum, the difference between the shopper and the detective is best characterized as follows: on the one hand, the shopper makes a mistake in making an inference guided by the formal end of practical reason from a general, prima-facie attitude to a particular, all-out one. On the other hand the detective makes a mistake in making an inference guided by the formal end of theoretical reason from a particular (prima-facie) attitude to a general (all-out) one. Because the move from the general to the particular in the practical realm is non-trivial, all-out practical attitudes are always liable to mistakes of performance.

Now if all mistakes in practical reason were mistakes of performance, we would have an argument for the Humean Theory of Practical Reasons that could probably ground an argument for the Humean Theory of Motivation; after all, the best candidates for non-desire-based reasons are general in character. But this view is obviously false; Leo might be mistaken not only in placing the wrong fruits in the cart, but also in his general intention to see to it that there are grapes in the cart. But couldn't we generate an argument against BB moral cognitivism from the fact that the conclusion of practical reason must be particular in character? After all, moral beliefs are general in character and if one cannot act without making a non-trivial move from the general to the particular, one needs something beyond moral belief to be motivated to engage in any particular action. However this gives us no reason to reject the view that these beliefs can motivate one to act in a particular way in accordance to a general intention. The most we could rule out is that an action could be solely motivated by a moral general belief. This would not necessarily be a problem for the motivational cognitivist. Think for instance about a principle of beneficence such as:

(B) One ought to help others.

Now assume that an agent finds herself in a situation where she could help someone out of the subway (suppose there's a wide gap between the door and the platform). She can do this by either giving the passenger a hand, or by lying down, head inside the train and feet in the platform, so that the passenger could walk over her back in a mildly painful way (she has a

[51] This is admittedly just a sketch of an argument for these claims. I provide more detailed argument in Tenenbaum (forthcoming (*b*)).

strong back, and the passenger is pretty light). I assume that the latter way of helping is, albeit awkward, morally permissible. Thus the following is arguably a consequence of (B):

> The action of helping the passenger by lying between the platform and the train is prima-facie good.

Nonetheless one would not necessarily conclude that such an action was correct or justified.[52] In general, what this shows is that, even in the case of perfectly virtuous action, moral belief alone cannot explain *every single aspect* of the action. But of course no sane form of motivational cognitivism should be committed to the opposing view.

Many attempts have been made to use the notion of direction of fit to expose significant differences between beliefs and desires, or to reveal a deep dissimilarity between theoretical inquiry and intentional action. I have been arguing that direction of fit does not lend itself well to these purposes, and in particular, that it does not lend support to the Humean Theory of Motivation. Surprisingly, attempts to render the notion of direction of fit more precise suggest a picture of reason in which there is in fact a deep similarity between the realm of practical reason and intentional action on the one hand, and the realm of theoretical reason and belief on the other.[53] This is the picture of a natural home for motivational cognitivism, a view in which one employs the same sort of rational faculties, albeit in relation to two different formal ends, in theoretical and in practical reason. Of course, I do not want to let the pendulum swing to the opposite error and argue that my reconceived account of direction of fit can prove the truth of motivational cognitivism; what I intend to convey here is just the sense that this notion may furnish valuable materials for rendering the view more plausible and precise.

REFERENCES

Altham, J. E. J., 'The Legacy of Emotivism', in G. Macdonald and C. Wright (eds.), *Fact, Science and Morality* (Oxford: Blackwell, 1986), 275–88.

Anscombe, G. E. M., *Intention*, 2nd edn. (Ithaca, NY: Cornell University Press, 1963).

[52] Of course what counts as "justified" and "correct" in practical reason, if anything, is the subject of much dispute, but the motivational cognitivist is bound to accept that some actions are justified or correct and others not. No matter how she does it, a reasonable motivational cognitivist should allow that this action might not be justified or correct.

[53] This picture of practical reason and intentional action is developed in much more detail in Tenenbaum (forthcoming (*a*)).

Bratman, Michael, 'Intention, Belief, Practical, Theoretical', unpubl. MS.
Clark, Philip, 'Velleman's Autonomism', *Ethics*, 111 (2001), 580–93.
Davidson, Donald, 'Intending', in *Essays on Actions and Events* (New York: Oxford University Press, 1980).
Humberstone, I. L., 'Direction of Fit', *Mind: A Quarterly Review of Philosophy*, 101 (1992), 59–83.
Kant, Immanuel, *The Critique of Pure Reason*, tr. Paul Guyer and Allen Wood (Cambridge: Cambridge University Press, 1998a).
____ *Groundwork of the Metaphysics of Morals*, tr. Mary Gregor (Cambridge, Cambridge University Press, 1998b).
Korsgaard, Christine, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996).
McDowell, John, 'Virtue and Reason', in *Mind, Value, Reality* (Cambridge, Mass.: Harvard University Press, 1998).
Nagel, Thomas, *The Possibility of Altruism* (Princeton: Princeton University Press, 1970)
Platts, Mark, *Ways of Meaning* (London: Routledge & Kegan Paul, 1979).
Railton, Peter, 'On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action', in G. Cullity and B. Gaut, *Ethics and Practical Reason* (Oxford: Oxford University Press, 1997).
Schueler, G. F., 'Pro-Attitudes and Direction of Fit', *Mind*, 100 (1991), 277–81.
____ *Desire: Its Role in Practical Reason and the Explanation of Action* (Cambridge, Mass.: MIT Press, 1995).
____ *Reasons and Purposes: Human Rationality and the Teleological Explanation of Action* (New York: Oxford University Press, 2003).
Shah, Nishi, 'How the Truth Governs Belief ', *Philosophical Review*, 112 (2003), 447–82.
Shaver, Rob, 'Sidgwick on Moral Motivation', unpub. MS.
Smith, Michael, 'The Humean Theory of Motivation', *Mind*, 96 (1987), 36–61.
____ *The Moral Problem* (Oxford: Blackwell, 1994).
____ 'Is there a Nexus between Reason and Rationality?', in Sergio Tenenbaum (ed.), *New Essays in Moral Psychology* (Amsterdam: Rodopi, forthcoming)
Sobel, David, and Copp, David, 'Against Direction of Fit Accounts of Belief and Desire', *Analysis*, 61 (2001), 44–53.
Tenenbaum, Sergio, 'The Judgment of a Weak Will', *Philosophy and Phenomenological Research*, 49 (1999), 875–911.
____ *Appearances of the Good* (Cambridge: Cambridge University Press, forthcoming (*a*)).
____ 'The Conclusion of Practical Reason', in Tenenbaum (ed.), *New Essays in Moral Psychology* (Amsterdam: Rodopi, forthcoming (*b*)).
Van Fraassen, Bas C., *The Scientific Image* (Oxford: Oxford University Press, 1980).
Velleman, David, *The Possibility of Practical Reason* (New York: Oxford University Press, 2000a).
____ (2000b) 'The Guise of The Good', in Velleman (2000*a*).
____ (2000c) 'The Possibility of Practical Reason', in Velleman (2000*a*).

Velleman, David (2000d) 'What Happens When Someone Acts', in Velleman (2000*a*).

Wedgwood, Ralph, 'The Aim of Belief', *Noûs-Supplement: Philosophical Perspectives*, 16 (2002), 267–97.

Williams, Bernard, 'Deciding to Believe', in *Problems of the Self* (Cambridge: Cambridge University Press, 1973), 136–51.

Zangwill, Nick, 'Direction of Fit and Normative Functionalism', *Philosophical Studies*, 91 (1998), 173–203.