**THE NATURE OF MOTIVATION**

 **(AND WHY IT MATTERS LESS TO ETHICS THAN ONE MIGHT THINK)**

ROBERT NOGGLE

Central Michigan University

**This is an accepted author manuscript. The final version is published in *Philosophical Studies,* Volume 87 (1997): 87-111.**

Recently there has been a resurgence of interest in the debate between the Humeans and Anti-Humeans about the role of reason and belief in motivation. This debate is over whether the springs of action consist of something like an argument whose premisses and conclusions have truth values in the way beliefs do. Three distinct versions of this debate have emerged, each of which embodies a fairly distinct conception of what exactly is at issue between the Humean and Anti-Humean. One version concerns the power of practical reason to motivate, where `motivate' is meant as a psychological term referring to the cause (or one of the causes) of action. Another concerns whether or not desires in and of themselves produce reasons to act, where `reason' is a term of rational justification rather than causal explanation. A third concerns whether motivation is a process which requires mental states other than beliefs.[[1]](#endnote-1)

It is the third version of the debate that interests me. At issue here is whether beliefs are motivational forces in and of themselves, or whether some mental state other than a belief (such as a desire) is necessary for motivation. For present purposes, we can gloss `motivation' in the following very rough way. Where Φ is an intentional action, for S to be motivated to Φ is for S to possess a mental state which tends to cause S to try to Φ (if S believes that Φ-ing is an available option).[[2]](#endnote-2) Since the notions of belief and desire play such large roles in this debate, we need to be as clear as possible about what the difference between the two is supposed to be. While both belief and desire are propositional attitudes, beliefs are supposed to be mental analogs of assertions. To believe that **P** is in a sense to mentally assert that **P**. Desires, on the other hand, are not supposed to be mental analogs of assertions. To desire that **P** is not to assert that **P**, mentally or otherwise. So beliefs, let us say, are assertoric propositional states, and desires are not. A similar idea often found in the literature is that desires and beliefs are propositional attitudes with opposite "directions of fit" with the world. Beliefs are supposed to be states that "aim" to conform to the world, while desires are supposed to be states that "aim" at getting the world to conform to them[[3]](#endnote-3). I will have much more to say about direction of fit below.

The debate over the respective roles of desire and belief in motivation is especially interesting for two reasons. First, it has very serious implications for folk psychology. The Anti-Humean claims that it is not necessary to postulate the existence of desires to explain motivation. She claims that all the explanatory work can be done by the notion of a motivating belief, and that ordinary talk of desires is underwritten by facts about beliefs rather than facts about some mental entity distinct from beliefs. If this is correct, then it seems to call into question our folk psychology--and the philosophy of mind and action theory based on it--which depends on the belief-desire mode of action explanation. I do not want to discuss here whether folk psychology will be replaced or eliminated when we get a complete scientific psychology. The point is rather that if desire is not really necessary for action-explanation, folk psychology as we now know it may be in considerable trouble long before we get a complete scientific psychology.

Second, this version of the debate is interesting because of its importance for meta-ethics. An anti-Humean thesis--that desire is not necessary for motivation--can be used to make plausible a kind of internalist moral realism. This theory unites two attractive claims. The first is the internalist claim that moral judgments are inherently motivating. The second is the realist claim that moral judgements assert propositions which have truth values in the same way that the contents of beliefs do.

Internalist moral realism is a sort of Holy Grail of meta-ethics. It offers us all we ever wanted from morality. The internalist claim gives morality the psychological "oomph" it needs to motivate action by itself, rather than having to hitch a motivational ride on pre- or non-moral motives. The realist thesis makes morality what it seems to be: a discourse about facts--moral facts--which we can discover, about which we can disagree, and of which we can often convince each other. Non-realist versions of moral internalism--such as emotivism and prescripitivism--build motivational force into moral judgments by claiming that they consist of expressions of desires, emotions, or prescriptions--none of which assert propositions. Other versions of moral realism hold that moral judgments do report moral belief--and thus do assert propositions--but often deny that moral judgements are inherently motivating, since a desire to be moral is needed before one will act on one's moral beliefs.

Internalist moral realism has been attractive to those ethicists who have wanted to hold that moral judgments are the apprehension of moral facts, and who at the same time do not want to see moral motivation as psychologically "optional" in the way that externalism would make it seem. Because its twin theses make morality everything that many of us thought it was before we became philosophers, internalist moral realism has the rather odd honor of being a theory that many of its opponents wish were true. In order to have both internalism about moral motivation and moral judgements which assert moral propositions, one must apparently hold that moral beliefs can motivate. To say that, one must apparently take the Anti-Humean position that beliefs can motivate in and of themselves without any help from desires.

**THREE THEORIES OF MOTIVATION**

**Radical Conativism**

There are three main philosophical theories about the roles of belief and desire in motivation. The simplest is Radical Conativism. It holds that desires are necessary for motivation and that belief has no effect on motivation whatsoever[[4]](#endnote-4). In essence, it claims that beliefs are "motivationally inert," since they make no causal contribution to motivation.

Radical conativism is just false, at least for human beings. This, I take it, is not particularly surprising. But it is important to see just why it is false. Belief clearly does affect our motivation. The entire enterprise of practical reason, as well as much of the folk psychology with which we quite successfully predict each other's behavior, depends on the fact that our motivations can change in response to changes in our beliefs. Certainly if radical conativism were true, we could still reason about what action might bring about the states we desire. But it is clear that we do more than reason about such things. When we acquire new beliefs about how to best fulfill some desire, almost invariably we come to have some desire toward the means, even if it is only a weak desire. The very ubiquity of this phenomenon makes it easy to overlook. It occurs literally hundreds, perhaps thousands, of times a day to each of us. Any time we come to desire the means to ends we desire, the phenomenon has occurred. If radical conativism were true this phenomenon would be utterly inexplicable, even miraculous. Thus we want **X** because we want **Y** and believe that **X** is a means to **Y**; we want **Y** because we want **Z** and believe that **Y** is a means to **Z**; and so on. Such a structure requires that desires be sensitive to means-end beliefs. It's a crucial fact about us that our motivation has this hierarchical structure and is quite responsive to our means-end beliefs. It is so crucial that any plausible theory of motivation must account for it. Radical conativism cannot do this, and that is why it is implausible.

**Cognitivism**

Diametrically opposed to radical conativism is cognitivism, which claims that belief alone produces motivation without the need for desires[[5]](#endnote-5). Cognitivism--as I conceive it--claims that a certain kind of belief (I'll call it a "motivating belief") is sufficient to produce motivation. The content of such a belief is that some state of affairs (often one involving the agent's action) prima facie ought to be brought about. Thus the content of a motivating belief is complex: it consists of a property asserted of a proposition. It may be a general "prima facie ought-to-be-brought-about" property, or it may be one of a great many properties. In the latter case, we can define a general prima-facie-ought-to-be-brought-about property as the disjunction of those properties such that a person must believe that the state of affairs has one of them before she can be motivated to bring about that state of affairs. It is an empirical question what properties a person must believe a state of affairs has before she will be motivated to try to bring it about. So it is an empirical issue what properties one must believe a state of affairs has before she will attribute the general prima-facie-ought property to it, or what properties we should include in the disjunctive summary prima-facie-ought property.

In particular, the "ought property" is not a specifically moral ought. Indeed, it should be thought of as a marker of "bare normativity"‑‑the thinnest possible notion of ought. Actually, the best rendition into English of the ought predicate would be something like: "I have prima facie reason to bring it about that **P**" It is this notion of ought--according to the cognitivist--that is such that believing it to apply to a state of affairs is a necessary condition for being motivated to bring about that state of affairs[[6]](#endnote-6). In addition to the claim that motivating states are beliefs about what ought to be brought about, a fully worked out cognitivist theory of human motivation will include generalizations meant to capture empirical facts about when people do believe that a state of affairs has this thin prima-facie-ought property, that is, what the necessary cognitive conditions for motivation are. The basic claim of cognitivism, though, is that a person must attribute the prima-facie-ought-to-be-brought-about property to a state of affairs if she is to be motivated to bring about that state of affairs.

If cognitivism is to be plausible, it will have to claim that certain motivating beliefs--what we might call practical axioms--are basic in the sense that they are not derived logically from or produced causally by any other motivating beliefs. The content of these basic motivating beliefs (and in particular their moral content, if any) is an empirical matter left open by cognitivism. Basic motivating beliefs can combine with non-motivating beliefs to produce new motivating beliefs which can in turn produce still other motivating beliefs recursively. Presumably beliefs with contents of the form "the state of affairs **Q** promotes state of affairs **P**," where **P** is a state of affairs to which some other motivating belief attributes the ought-to-be-brought-about property, will (ceteris paribus) lead the person to attribute the ought property to **Q**. So according to cognitivism, the motivational structure resembles a practical geometry, based on theorems derived from practical axioms. Roughly then, cognitivism holds that if **S** has a motivating belief that **P** prima facie ought to be brought about and believes that **Q** promotes **P**, then **S** will, ceteris paribus, have a motivating belief that **Q** prima facie ought to be brought about. Because it has this structure, cognitivism has no difficulty accounting for the predominant instrumental rationality and hierarchical structure of human motivation.

**Moderate Conativism**

Moderate conativism holds that while desire is causally necessary for motivation, beliefs can also play a causal role in motivation. In particular, moderate conativism holds that some desires are jointly caused by beliefs and other desires. On the most plausible versions of moderate conativism, the relation that normally holds between a desire on the one hand, and the belief and desire that caused it on the other, is something like the following: If **S** desires that **P** and believes that **Q** promotes **P**, then **S** will, ceteris paribus, desire that **Q**. Notice that this schema is very similar to the one that cognitivism posits. Because it embodies this schema, moderate conativism can account for the predominant instrumental rationality and hierarchical character of human motivation just as well as cognitivism can.

Plausible forms of moderate conativism do not hold that all desires are jointly caused by beliefs and other desires. The most plausible versions hold that innate instincts and biological forces such as drives and homeostatic mechanisms produce or constitute basic desires that can then combine with beliefs and produce non-basic desires recursively. What basic desires there are (and what their moral content is, if any) is an empirical question that moderate conativism leaves open[[7]](#endnote-7).

**MOTIVATIONAL AKRASIA**

I noted above that human motivation is largely instrumentally rational, given our means-end beliefs. But interwoven with this fabric of subjectively rational motivation we find pockets of subjective irrationality, which I shall call "motivational akrasia." Let the term `beneficial' have the following technical use: a person believes something is beneficial just in case she believes that it promotes some state that she desires. Motivational akrasia occurs when a person's motivations do not conform completely to her beliefs about what is beneficial. There are two forms of motivational akrasia. In the strong form, a person lacks any desire for something she believes is beneficial. In the weaker form, a person's strongest desire is not for what she believes is most beneficial, although she does have a desire (of some strength) for everything she believes beneficial. I take it to be uncontroversial that both forms of motivational akrasia exist. Unfortunately few things are more familiar than the situation in which a person believes that he is better off if he does something--exercise, work, change the cat's litter--and yet cannot "get motivated" or get sufficiently motivated to actually do it. Motivational akrasia is as important a feature of human motivation as its hierarchical structure and predominant means-end rationality. And thus any plausible theory of motivation must allow for its possibility.

Moderate conativism may seem to rule out the existence of motivational akrasia by fiat because it claims that (non-basic) desires are caused by beliefs and other desires. Because it gives motivational force to beliefs and desires, it may not seem to leave enough room for motivational akrasia--that is, for cases in which a person's beliefs and desires do not motivate. But notice that moderate conativism requires only that each non-basic desire be produced partly by a relevant belief (that is, a belief the content of which bears logically on the fulfilling of that desire). It does not require that each relevant belief produce a desire. Thus some relevant beliefs may produce desires and some may not. Why, how, and when it happens that a desire and a relevant belief do not produce a desire are interesting questions for moderate conativism, but they do not seem unanswerable. Perhaps there is a size limitation on the working memory which holds the beliefs that are possible inputs for desire generation. Sometimes relevant beliefs might not make it into this working memory for "mechanical" reasons and so do not become the cognitive input for the formation of any desires[[8]](#endnote-8).

Moderate conativism also allows plenty of room for weak motivational akrasia. Moderate conativism does not require one's desires to be ordered in strength according to one's beliefs about the relative beneficialness of their objects. If a person believes that action a is very beneficial and believes that b is only moderately beneficial, nothing in moderate conativism requires her to desire to perform a more than she desires to perform b. So while the existence of weak motivational akrasia is an important fact about us--one for which any plausible theory of motivation must allow--it presents no obvious threat to moderate conativism.

What about cognitivism? It is tempting to think that the fatal flaw in cognitivism is that it cannot account for motivational akrasia. It seems to posit too close a link between belief and action to allow room for it[[9]](#endnote-9). But cognitivism has ample resources for dealing with motivational akrasia. Cognitivism does not require that every relevant practical belief produce motivation. Rather, it requires that every motivation be produced by some practical belief, and leaves it open that some practical beliefs might not (e.g., for "mechanical" reasons, perhaps having to do with the size of working memory) become inputs to the processes that produce motivation.

And cognitivism, like conativism, is not a thesis about the strengths of motivations, so weak motivational akrasia does not present any obvious problems for it. In both moderate conativism and cognitivism the explanation of weak motivational akrasia will presumably involve emotional states, salience effects, and various "mechanical" features. Thus the agent's current emotional state might dampen or heighten certain motivations out of proportion to the strength the motivation "ought" to have based on the expected utility theorem (or some other principle of normative decision theory)[[10]](#endnote-10). Whatever the causes of weak motivational akrasia are, both moderate conativism and cognitivism are compatible with cases in which a belief that **a** is more beneficial than **b** occurs with a desire for **b** that is stronger than the desire for **a**.

It appears, then, that both conativism and cognitivism can allow for motivational akrasia in either of its forms. And we saw that each can account for the hierarchical character and predominant instrumental rationality of human motivation. These two facts about human motivation seem so crucial that any plausible theory of motivation must account for them. And both conativism and cognitivism are initially plausible insofar as they can account for the rather striking fact that human motivation consists both of a high degree of instrumental rationality and occasional (but occasionally very problematic for its victim) motivational akrasia.

I turn next to the debate between cognitivism and moderate conativism. Since radical conativism is a non-starter about which I have nothing further to say, I will drop the `moderate' from `moderate conativism' use the term `conativism' unqualified to refer to that theory in what follows.

**THE COGNITIVIST'S CHALLENGE**

Several philosophers have pointed out that our conceptual scheme seems to require the notion of desire in action explanations. That is, the term `desire' seems to be synonymous (or nearly so) with the term `motivation', so that it is analytic that if a person did something voluntarily then she desired to do it. Now the cognitivist might agree it is conceptually necessary to say that agents perform actions because they desire to do so. But she will (indeed, she must) deny that this conceptually necessary talk of desires is underwritten by the existence of mental entities separate and distinct from beliefs. Instead, she will claim that the truth of commonsense talk of desires is grounded in facts about beliefs and whether they had motivational force on a given occasion. So while our conceptual scheme may force us to use desire-talk in action explanations, the cognitivist will claim that this is just short-hand for statements about which beliefs possess motivational force at a given time.

In this way the cognitivist might accept the use of desire-talk while maintaining that the postulation of a desire that is a special mental state other than a belief is not necessary. In other words, the cognitivist offers a deflationary account of commonsense talk of desires whereby it is completely underwritten by beliefs. (Alternately, she may making a reductionist claim that desires can be reduced to--and perhaps eliminated in favor of--beliefs.)

Conativism, however, requires the existence of "real" desires--separate and distinct from beliefs. But if--as the cognitivist suggests--there is no explanatory need for them, then there is no reason to accept conativism. Thus the cognitivist offers the following challenge to the conativist: "Show me why we need this `robust' desire talk, necessarily underwritten by motivating states distinct from beliefs." If the conativist cannot do this, then there is apparently no reason to postulate the existence of "real" desires to back our desire-talk, and thus no reason to be a conativist. Thus the cognitivist challenges the conativist to show what is to be gained by going beyond the cognitivist's deflationary (or eliminativist) reading of desire-talk[[11]](#endnote-11).

The cognitivist's challenge is a challenge to the entire system of belief-desire explanation[[12]](#endnote-12). It challenges it on grounds of economy: why explain action in terms of two distinct types of mental states when one will do? Given the popularity of the belief-desire mode of action explanation among both philosophers and laypeople, one might expect that replying to the cognitivist challenge would be easy. After all, since we have been using beliefs and desires to explain action all along, it might seem that we must have been doing so with good reason. But, as I will try to show below, it is actually very difficult to reply to the cognitivist's challenge.

**ATTEMPTS TO MEET THE COGNITIVIST'S CHALLENGE**

**Arguments Based on Goals**

The conativist might attempt to meet the cognitivist's challenge by claiming that desires are necessary to account for the goal-directedness of motivation and action. Motivation is inherently a notion involving goals: one is motivated toward some object. The conativist might argue that this fact is incompatible with the claim that motivations are beliefs. Desire, she might claim, is necessary to capture the fact that motivation involves goals. If she can show that a belief could not embody or express a goal in the right way to account for the goal-directedness of motivation, then it would appear that she will have met the cognitivist's challenge.

It might seem fairly obvious that motivations--especially basic motivations--could not be beliefs because they seem to be purely goal-directed states. That is, motivations like hunger or the sex drive just do not seem to be very belief-like. Instead, they seem to be pure impulsions, things much better captured by the notion of desire than that of belief. In order to evaluate this line of thought, we will have to distinguish between two general kinds of basic motivations. Basic motivating states may be roughly divided into the relatively static and the relatively dynamic, depending on the extent to which their existence or their strength is stable over time. More or less static basic motivations would include a drive for self‑preservation, the avoidance of pain, and so forth. While they are not constantly expressed, they do seem to be the kinds of motives that do not go away or vary in strength over time. They can be captured in the cognitivist vocabulary as basic practical axioms: "Pain ought prima facie to be avoided," "Life ought to be preserved," and so on. Their content is always relatively fixed, even though they do not always play a role in motivation. More dynamic motives are things like hunger and thirst. They typically result from (or consist of) homeostatic mechanisms, and thus their strength (or perhaps even their existence) varies over time depending on the state of the homeostatic system. These dynamic motives can be seen from the cognitivist point of view as complexes involving information detectors plus practical axioms. The information detectors render declarative judgments about the condition of the relevant homeostatic system. And the practical axiom states that if the homeostatic conditions go out of balance, then certain actions ought to be performed. Hunger, for example, would consist of a practical axiom to the effect that if one's blood-sugar level falls, then one ought to eat, along with an indicator of the current blood-sugar level. So again, it seems that the cognitivist can account for and describe the same motivational phenomena that the conativist can, but without postulating desires.

Yet the conativist might still think that there is something special about goals that the notion of desire captures and the notion of belief does not. If, as I have suggested, the basic difference between beliefs and desires is that beliefs are assertoric and desires are not, then this claim becomes the claim that only a non‑assertoric state could express a goal in the right way to account for motivation. The conativist might be tempted to say that goal states cannot be beliefs because the contents of beliefs are expressed by indicative sentences whereas a goal must be expressed by something other than an indicative sentence (e.g., an imperative). If this is right, then it would be true that the concept of a goal is better captured by the notion of desire than that of belief.

The claim that only desires express goals in the right way to account for action might seem to follow directly from the fact that beliefs are assertoric and desires are not. A desire is nothing like the assertion that something is the case. But it is like the expression of a goal. To say that I desire **X** is very nearly the same as saying that my goal is **X**. Beliefs, it would seem, do not express goals; they merely represent (or aspire to represent) facts. And a statement of a fact would seem to be a completely different logical critter from an expression of a goal. The conativist might want to latch onto this apparently fundamental difference and claim that only states that express goals could be the causes of action.

Though this line of thought is quite tempting, it does not, I think, bear close examination. For motivating beliefs also involve goals, even though they are assertoric. A motivating belief attributes the ought property to a state of affairs toward which the believer is motivated. This state of affairs, then, is the goal. So it is simply false that motivating beliefs do not involve goals. Of course it is true that the goal plays a syntactic role in the sentence expressing the content of the motivating belief which is far different from the syntactic role it plays in the sentence expressing the content of a desire. The statement of the goal all by itself expresses the content of a desire. But the statement of the goal is embedded in the sentence that expresses the content of a motivating belief. It is the state to which that sentence attributes the ought property.

Yet it is difficult to see why this should make a difference. Suppose it is true that a state must express a goal to account for action. Desires express goals in a very straightforward manner. But why doesn't attributing the ought property to some state count as expressing a goal? The cognitivist, of course, will argue that it does. She will claim that attributing the ought property to a state of affairs expresses the goal that this state of affairs come about. She will further claim that this kind of goal expression can account for the goal-directedness of motivation.

In order to refute the cognitivist here, the conativist will have to claim that only states that express goals as something like imperative sentences could cause actions. But this claim seems false. To see why, we have only to look at computers. Many computer architectures are goal driven. The statements that express the goals in such an architecture are often imperatives. But nothing prevents them from being indicative. All that is required is that the architecture be arranged so as to attempt to bring it about that **Q** when it gets a statement indicating that **Q** is the goal. That statement can be an indicative, something like "the goal is now **Q**"; or it can be an imperative, something like "Let it be that **Q**." All that is required is that the architecture be designed to parse the goal expression, extract **Q** as the goal state, and then initiate activity that will (attempt to) bring about **Q**. There is no difficulty--either conceptual or practical--with setting up a system to use goals expressed either as imperatives or as embedded within indicatives. And I can see no a priori reason why the human motivational architecture must operate on goal statements expressed by imperative sentences. And so while it does seem that motivation involves goals, it does not appear that this fact helps the conativist meet the cognitivist's challenge.

**Direction of Fit Arguments**

Much of the recent literature on the controversy between the Humeans and Anti-Humeans has appealed to the notion of "direction of fit" as a way of meeting the cognitivist's challenge. The idea is that beliefs and desires have different "success conditions" and that only states with certain success conditions could account for motivation. Beliefs are said to be successful only if they accurately represent the way the world already is. Desires, on the other hand, are successful only if they lead to the world conforming to the content of the desire. Thus a belief that **P** is successful if it is already true that **P**, while a desire that **P** is successful if **P** eventually comes about. Thus desires are "active" in a certain way: they are successful only if the world changes to conform to them. Beliefs are "passive:" they are successful only if they (merely) reflect the current state of the world. States that are passive in the way that beliefs are supposed to be are said to have a "direction of fit" opposite from that of states that are active in the way that desires are supposed to be.

Conativists who have appealed to direction of fit generally make two claims: first, that desires and beliefs have a distinct direction of fit with the world, and second, that only states with the "active" direction of fit of desires could cause actions. Thus the argument is supposed to go like this:

Premise One: Desires have a distinctly active "world to state" direction of fit.

Premise Two: Only states with this active world to state direction of fit could cause action.

Conclusion: Desires are necessary for the causation of actions.

Michael Smith offers the following version of the argument[[13]](#endnote-13). He claims that states which motivate action must both represent the state of affairs to be brought about and be compatible with a belief that the state does not now obtain. Since beliefs are in general incompatible with beliefs that deny that their contents obtain, beliefs cannot fill this role. Desires, however, because of their direction of fit, are representations of states that are to be brought about but which to not yet obtain. Thus, Smith concludes, only desires could account for motivation.

Once we see what the content of a motivating belief is, it is pretty easy to see where this argument goes awry. A belief that a state of affairs ought to be brought about certainly is incompatible with another belief denying that it ought to be brought about. But it is quite compatible with a belief to the effect that the state of affairs that ought to be brought about does not now obtain. In short, both a desire that **P** and a belief that **P** ought to be brought about are compatible with a belief that not-**P**. Thus Smith's argument fails to show that only a desire could have the "active" direction of fit said to be necessary for motivation[[14]](#endnote-14).

Smith's version of the direction of fit argument is only one of several in the literature. Some versions appeal to a normative notion of direction of fit. The norm for a belief is for it to reflect what is already true; that is, the belief must reflect the world. But the norm for a desire is for the world to reflect the content of the desire; that is, the world must be made to match the desire. Thus it is a criticism of a belief that it is not true. The belief in some sense "fails" if it is false. But it is not a proper criticism of a desire that the state of affairs that is its object has not come about. Rather the norm for desire dictates that it is the world that ought to conform to the desire. The desire does not "fail" if the state of affairs that is its object does not obtain. Rather it is the world that in some sense "fails." Can such a normative notion of direction of fit be used in the argument above? Is it true that only desires are states with which the world "must" fit? Is it true that only such states can cause actions?

First, I have my doubts as to whether normative criteria for a mental state could determine whether or not it could cause action. Indeed, it is difficult to see why any normative criteria governing a mental state should have anything at all to do with what sorts of causal relations it can partake in. The causal powers a state has are determined by how it is connected to other states. Surely a system could be hooked up causally so that states with the passive state-to-world direction of fit (such as beliefs about what states of affairs ought to be brought about) produce motivation. And so I do not see why explaining the notion of direction of fit in terms of norms would make Premise Two true. But I leave that worry aside.

Instead, I want to question the claim that, on this reading, the direction of fit of desire is always distinct from that of belief (Premise One). It seems to me that the very same normative criteria that apply to desires also apply to motivating beliefs. It is true that a motivating belief, like any other belief, "fails" if its content is not true. But recall that the content of a motivating belief is complex: it is an ascription of the ought property to some state of affairs. Now if that ascription is false, then, of course, the belief has "failed." But if the state of affairs to which the ought property is ascribed does not obtain, then it seems that it is the world which has failed and should be made to change. That's pretty much the meaning of `ought', after all. So if I have a belief that **P** ought to come about, the fact that not-**P** does not mean that the belief has failed to fit the world, but rather that the world has failed to fit the belief. If the world should be changed to fit a desire, then surely it should be changed to fit an ought belief. So it seems that the motivating belief bears the same normative relation to the state of affairs to which the ought property is ascribed as the desire bears to the state of affairs that is its object. Thus while the motivating belief has the "passive" belief-like direction of fit with regard to its whole content, it has the "active" desire-like direction of fit with regard to the state of affairs to which it ascribes the ought property. So arguing for conativism on the basis of this construal of the notion of direction of fit does not seem promising. Because both desires and motivating beliefs have the same direction of fit with regard to the (non-actual) state of affairs that is to be brought about, Premise One turns out to be false.

Some conativists have suggested that direction of fit is not, or not merely, a normative notion, but a causal or functional one. The idea is that desires but not beliefs have a distinct causal power to get the world to conform to their contents. Beliefs, on the other hand, are caused by the world to have their contents reflect the world. Beliefs function as representations, while desires function to change the world to fit their contents. Given this reading of the direction of fit metaphor, is Premise Two true? That is, is it reasonable to think that only desires have the direction of fit needed to bring about action?

I don't think so. It still seems false that only desires have the "active" direction of fit. Consider the belief that **P** promotes **Q**. In a person who desires **Q**, this belief helps produce motivation to go after **P**; that is one of its causal properties, part of its function. Means-end beliefs, in creatures who desire the end, function to produce actions directed toward the means. That is why we call means-end knowledge practical knowledge: it is supposed to play a role in our actions. So again it appears that desires are not alone in having the active state-to-world direction of fit.

In a recent article, Lloyd Humberstone develops a very sophisticated account of direction of fit[[15]](#endnote-15). His idea is that the concept of belief involves a "background intention" not to believe what is false. This background intention is, he says, "constitutive" of the notion of belief: if one does not intend to cease to have an attitude toward a proposition if it is false, then that attitude is not belief. Humberstone postulates a similar background intention for desire: one intends that, if she desires that **P**, then **P**. He spells out these conditional background intentions more formally as:

Intend (~**B**(**P**)/~**P**) ("one intends that she not believe that **P**, given that not‑**P**")

Intend (**P**/**D**(**P**)) ("one intends that **P**, given that she desires that **P**")

Humberstone suggests a possible‑world semantics for these intentions: Of the worlds in which not‑**P**, we prefer those in which we do not believe that **P**. Of the worlds in which we desire that **P**, we prefer those in which **P** obtains[[16]](#endnote-16). So for belief, one starts with worlds in which not‑**P** and "looks for" those in which one does not believe that **P**. In this way, one is in some sense trying to get the belief to conform to the world. One is selecting among worlds based on whether or not the belief is present in them. With desire one begins with those worlds where one has the desire, and "looks for" the ones in which the object of the desire obtains. In this way one in some sense is trying to get the world to fit the desire. Humberstone thinks the key to the direction metaphor is this holding the "given" condition constant while looking for those worlds where the other condition obtains.

 For the sake of argument, let's grant that this formulation makes the notion of direction of fit less mysterious and more precise. But does it help the conativist? Remember that to do so, it must make states with the active desire‑like direction of fit necessary to account for action, and give only desires that direction of fit. Suppose Humberstone is right that certain background intentions account for direction of fit. Any argument for the conativist claim (Premise One in the argument above) that certain directions of fit are "proprietary" to certain kinds of attitudes would have to claim that only certain background intentions exist. The conativist would have to show that those background intentions which would give an active desire-like direction of fit to beliefs do not exist. It is difficult to see how such an argument would go.

Moreover, there is good reason to think that there are background intentions which would in fact grant the active desire‑like direction of fit to some beliefs. Presumably most of us have the following background intention:

Intend [**P**/**B**(**O**(**P**))] ("one intends that **P**, given that she believes that **P** ought to be

 brought about")

Surely we "intend"--in Humberstone's sense of that word--that **P** should come about given that we have a belief that **P** ought to come about. This is what `ought' means: if we believe that some state of affairs ought to come about, then we prefer that it does. Or, to use Humberstone's semantics, among all the worlds in which we believe that **P** ought to obtain, we prefer those in which **P** does obtain. Surely this is constitutive of believing that something ought to be the case. And if so, then on Humberstone's scheme, beliefs about what ought to be the case have the desire‑like direction of fit. If this is right, then even on Humberstone's ingenious proposal, it will not work out that only desires have the direction of fit (supposedly) needed to account for action. Ought beliefs will also have it, and since the motivating beliefs the cognitivist postulates are a species of ought beliefs, they too will have the desire-like direction of fit.

Judging from what we have seen so far, conativist arguments based on direction of fit do not seem promising. The metaphor itself is rather obscure even on a sophisticated reading such as that provided by Humberstone. Even more obscure is the conativist claim that only states with a certain direction of fit could cause action. And it is very difficult to see how any reading of it will give desires a direction of fit that motivating beliefs lack and at the same time make it turn out that only states with that direction of fit can produce motivation[[17]](#endnote-17).

**"WE HAVE MET THE ENEMY, AND HE IS US"**

So far things to not look promising for the conativist. Neither an appeal to goals nor an appeals to direction of fit seems to help her meet the cognitivist's challenge to capture some fact using desires that she can capture without them. I think that there are deep reasons for this. Indeed, I think that cognitivism will be able to say anything conativism that says. This is because any conativist statement involving the term `desire' seems to have a cognitivist "translation" devoid of the term `desire'. In general, any conativist statement involving a desire that **P** will be translatable into a cognitivist statement involving a motivating belief that **P** ought to be brought about.

Although conativism and cognitivism posit two apparently different kinds of states--desire and motivating belief--we can rewrite conativist laws that involve desires with simple propositional objects as laws involving a special kind of belief with complex propositional objects. Moreover, the cognitivist and the conativist will agree to a large extent about what content to attribute to these mental states. For any given motivational state, the cognitivist will call it a belief that **P** ought to be brought about, while the conativist will call it a desire that **P**. But--and this is the crucial part--**P** is the same proposition for each. Thus two field psychologists--one subscribing to cognitivism and the other conativism--may differ in labelling a subject's mental state as a desire that **P** or a belief that **P** ought to be brought about. Yet they will generally agree on what proposition **P** is. Moreover, the stories each tells about the interactions between the motivational states and other mental states will be the same. As we have seen, both theories can account for the mixture of predominant means-end rationality and occasional motivational akrasia that characterizes the human condition. And both seem equally compatible with whatever psychology might tell us about the most basic human motivations.

The isomorphism between the two sets of laws suggests that they are mere notational variants of the same theory. We can treat motivational states equally well as either motivating beliefs or desires. All the theoretical and empirical implications of the motivating belief‑containing and desire‑containing theories are the same. Because of this, they certainly seem to be mere notational variants of the same theory.

This rather surprising result is partly a consequence of embracing functionalism about propositional attitudes. For functionalism says that we must characterize mental state types by their causal roles. Since conativism and cognitivism both postulate motivating state types which fill isomorphous causal roles, functionalism apparently tells us that the theories are mere notational variants of each other. The only difference is that one notation characterizes motivational states as attributing an ought property to a goal state, whereas the other characterizes them as a special kind of attitude toward a goal state. But since the asserting of the ought property plays the same functional role within cognitivism that the special attitude plays in conativism, functionalism tells us that these two apparently different things actually turn out to be the same thing. This results in part from abandoning type-type identity in the philosophy of mind: since there is no physical brain state type corresponding to a mental state type (i.e., the motivating state type), we cannot simply appeal to what is "really" in the head to determine how to individuate mental state types. Instead, we individuate propositional state types by their functional roles. But, as I have argued, the functional roles of desire and motivating belief appear to be the same. And so functionalism tells us that there is no real difference between a motivating belief and a desire.

It may look as though the claim that conativism and cognitivism are notational variants depends entirely on rejecting type-type identity in favor of functionalism. This is not so, however. I think that conativism and cognitivism would be notational variants even if functionalism were not true. To see why, let us assume for a second that type-type identity were true. Let us assume further (what may or may not be the same thing) that brain states have physical features which correspond to the syntactic structure of the sentences that express the propositional contents of the mental states to which they are identical. That is, let us assume that any brain state which instantiates a mental state has parts or features which represent the various parts of the sentence that expresses the propositional content of the mental state. Thus a brain state identical to the belief that the cat is on the mat would have a part or feature marking it as a belief, a part or feature referring to the cat, and a part or feature attributing "on-the-mat-ness" to the cat.

Of course the recent history of the philosophy of mind provides little reason to think that brain states instantiate mental states in this manner. Quite the contrary in fact. But this would seem to be the most favorable picture of the mind-brain relation to make good on the claim that conativism and cognitivism are not mere notational variants. For on this picture, it would seem that there could be a real--physical--difference between a belief and a desire, and a fact of the matter about which is necessary for motivation. For the content of a motivating belief is complex, consisting of a proposition describing a state of affairs, embedded within another proposition which asserts that this state of affairs ought to be brought about. The content of a desire is just a proposition describing some state of affairs. Since the syntactic structure of the sentences that express the contents of the two mental states differ--one being an atomic sentence and one being a sentence with another sentence embedded within it--it would seem that the internal story that cognitivism will tell will be much different from that which conativism will tell, regardless of any similar predictions that the two theories might make. It seems that cognitivism could "really" be true and conativism false if it turned out that there was some feature that motivating states have in virtue of which these states are assertoric.

But I don't see how there could be any physical feature that would tell us which theory of motivation is "really" correct, even if brain states were like sentences in the head. In a physical instantiation of a propositional attitude there is still nothing intrinsic to the physical state of affairs which determines which parts of it represent the attitude and which parts represent the sentence expressing the proposition. That is, in any instantiation of a propositional attitude, some physical features will determine which attitude is being instantiated while others will determine which proposition is the content of that attitude. But nothing intrinsic to the physical state of affairs itself determines which features are which. So there is no way to look at the physical instantiation of a motivational state and determine whether it is a belief that **P** ought to be brought about rather than a desire for the obtaining of **P**. For any feature putatively representing the ought property, it will be open for the conativist to say that this feature is in fact part of a complex feature that marks the desire attitude. So where the cognitivist sees a feature representing the ought property, the feature representing some proposition, and a feature marking the belief attitude, the conativist will see a feature representing the proposition and a complex feature marking the desire attitude. And there is no fact of the matter which forces us to individuate the properties of some physical instantiation of a mental state in one way or the other. Nothing forces us to say that a state has two features--one which instantiates the belief attitude and one the ought property--rather than saying that it has one (complex) feature that instantiates the desire attitude.

If we can attribute content to the system in either of two ways, one which makes a given feature part of the attitude and the other which makes it part of the propositional content, then it seems as though the only thing on which to base a choice between the two attributions is explanatory power. But it looks as though the plausible versions of the two theories have exactly the same explanatory power. The basic motivational facts--especially facts about our predominant instrumental rationality mixed with occasional motivational akrasia--can be accounted for by either of two theories. One is conativist, one is cognitivist.

If the debate between conativism and cognitivism is really just a debate over terminology, then which notation should we prefer? Both conativism and cognitivism might seem to do some violence to our commonsense concepts of belief and desire. Cognitivists often complain that conativists use the notion of desire to cover all sorts of motivations whereas our "real" commonsense notion of desire has the connotation of a biological or quasi-biological urge. Thus cognitivists sometimes claim that the common person might well speak of desires for food or sex, but she is unlikely to speak of a "desire" to be moral. I think that conativism may in fact do some violence to the ordinary ways of speaking. Yet this violence is not at all as great as the cognitivist might insist. For though it is perhaps uncommon to talk of a desire to be moral, it is certainly not unintelligible. On the other side, conativists might charge that the notion of belief is really not meant to do the work of motivation, especially in cases involving basic biological motivations. But again, there is nothing incoherent or unintelligible in the notion that there could be a motivating belief--a mental state that is assertoric but which also creates motivation. Indeed, cognitivists are fond of providing intelligible descriptions that have the feel of ordinary talk and which attribute motivating force to beliefs. While it is not clear that the ordinary folk usually (if ever) do talk this way, such talk is certainly intelligible. So while I think that both conativism and cognitivism do some violence to our ordinary notions of belief and desire, this violence it not great in either case, nor is it particularly clear that one language does more violence than the other.

But it is true that the origins of desires/motivating beliefs are somewhat different from those of "regular" beliefs. The genealogies of "normal," non-motivating beliefs are bound up with sense organs. But the genealogies of desires/motivating beliefs are linked to the mechanisms which govern homeostasis or which constitute various biological drives or instincts. While it might be possible to portray these as (very specialized) information detectors, they are quite different from the all-purpose detectors that give rise to normal beliefs. It might be useful to separate the kind of mental state derived ultimately from drives, instincts, and homeostatic mechanisms from the kind derived ultimately from the canonical sensory organs. One way of doing this is to abandon the language of cognitivism in favor of that of conativism. And certainly the terminology of conativism is closer to ordinary practice. Most of us think and talk in terms of desires rather than motivating beliefs, and folk psychology and decision theory are both usually formulated in that vocabulary.

So perhaps there are at least pragmatic reasons to go beyond cognitivism and add the term `desire' to our psychological vocabulary. And so perhaps conativism is a slightly more clear and useful theory even if it does not fulfill the cognitivist challenge. But this is a very hollow victory. Conativism wins it only because its notation is a bit more convenient, not because it uniquely captures the nature of motivation.

**EPILOGUE: MOTIVATION AND MORALITY**

One might think that we could use results from ethics to help us break the tie between conativism and cognitivism. That is, one might think that we could settle the impasse by determining whether our theory of practical reasoning commits us to the claim that practical judgements both have truth values and are motivational forces in and of themselves. In that case, cognitivism would have to be true. So we might attempt to settle questions about motivation after settling debates in practical philosophy. I for one am quite suspicious of any attempt to say anything substantive about the psychological details of motivation based on arguments in ethics or decision theory. Indeed, if what I say here is right, there is nothing substantive to say about the nature of motivation that can be said only by conativism or only by cognitivism anyway. Arguments from ethics can merely be used to show us why we ought to pick one or the other notation. And this is a quite legitimate project, though a rather less exciting one than using ethics to discover some deep fact about motivational psychology[[18]](#endnote-18).

What my suggestion rules out--if it is right--is the project of using some thesis about the conative or cognitive nature of motivation to argue for some thesis in meta-ethics. Many ethicists worry about the debate between cognitivism and conativism because a variety of internalist moral realism claims both that moral utterances have truth values and that moral statements can motivate on their own, without the need for a separate desire. If one could show that cognitivism was true, that version of internalist moral realism would look quite plausible. On the other hand, if conativism turned out to be true, it would seem that one would have to choose between internalism about motivation or the sort of moral realism that claims that moral judgements assert propositions; one could not have both. Thus it has appeared that the outcome of the debate over the nature of motivation could have a profound impact on meta-ethics. But if I am right, then facts about human motivation can be captured equally well with conativist or cognitivist language. And if that is true, then nothing about motivation either implies or rules out internalist moral realism. So it appears that the debates over internalism and moral realism will have to occur within moral theory proper and not simply appeal to the nature of motivation[[19]](#endnote-19).

 **NOTES**

Department of Philosophy, DM 337

Florida International University

Miami, FL 33199

1. For a useful overview of the current debate, see R. Jay Wallace, "How to Argue about Practical Reason" (Mind 99 [1990]: 355-85) and David Brink, Moral Realism and the Foundations of Ethics (Cambridge, Mass.: Cambridge University Press, 1989), chap. 3. For an example of the first sub-debate, see Alfred Mele, "Motivational Internalism: The Powers and Limits of Practical Reasoning" (Nous 24 [1990]: 417-36); for an example of the second, see Thomas Nagel, The Possibility of Altruism (Princeton: Princeton University Press, 1970); for an example of the third debate, see Michael Smith, "The Humean Theory of Motivation" (Mind 96 [1987]: 36-61). [↑](#endnote-ref-1)
2. I can offer here no account of intentional actions, but will just assume that we know them when we see them. For brevity I will omit the `intentional' from `intentional actions' in what follows. All of the actions to which I will refer are intentional and voluntary unless otherwise noted. [↑](#endnote-ref-2)
3. The distinction is originally found in G.E.M. Anscombe, Intention (Ithaca:Cornell University Press, 1969). See Michael Smith, "The Humean Theory of Motivation," and Dennis Stampe, "Defining Desire" in Joel Marks, ed., The Ways of Desire (Chicago: Precedent, 1986). See also Robert Gordon, The Structure of Emotions (Cambridge: Cambridge University Press, 1987). As we will see, some theorists adds that the contents of belief and desires are related causally to the world in opposite ways. A belief is caused by the state of affairs its content describes, while a desire tends to bring about the state of affairs that its content describes. See Dennis Stampe, "Defining Desire." [↑](#endnote-ref-3)
4. What we might call "Extreme Radical Conativism" might go beyond this and add that desire is a non-intentional entity. Radical Behaviorism, at least on certain reconstructions, would probably be an example of such a theory. [↑](#endnote-ref-4)
5. Some versions of cognitivism do in fact use the term `desire' but say that desires are really a species of belief. I will say more about this in a bit. [↑](#endnote-ref-5)
6. To make this really precise, we would have to relativize all this to actions believed to fall under certain descriptions. But I shall leave such complications aside here. [↑](#endnote-ref-6)
7. In addition to cognitivism, radical conativism, and moderate conativism, one could hold a mixed view of motivation according to which some motivations are caused by desires and others by beliefs. It has been suggested that the debate between the Humean and anti-Humean is best characterized as a debate between some pure form of conativism and a mixed theory which states that desires often do contribute to motivation but that on some occasions beliefs--especially moral beliefs--can motivate on their own. (Garret Deckel pressed this point in her comments on an earlier version of this paper at the Pacific Division APA. As she pointed out, this seems to be the position of Thomas Nagel in The Possibility of Altruism.) This way of characterizing the debate may well be truer to the positions of actual cognitivists than the way I am characterizing the debate. We will see in a moment, however, that the pure cognitivist--as I construe her position--can offer a challenge to ever postulating desires in explanations of motivation. That argument threatens both the pure forms of conativism and mixed views alike. [↑](#endnote-ref-7)
8. This suggestion is inspired by some of Christopher Cherniak's remarks in Minimal Rationality (Cambridge: Cambridge University Press, 1986). [↑](#endnote-ref-8)
9. See Michael Smith, "The Humean Theory of Motivation," for such a view. [↑](#endnote-ref-9)
10. I do not mean to advance any particular theory about what does in fact affect motivational strength. That is a question for empirical psychology. But in any case it is clear from the fact of weak motivational akrasia that the effective motivational strength of a motivational state depends on more than just the subjective probability that the means will secure the end and the strength of the motivation toward the end. No doubt it will depend on these things, but it will also depend on a host of mechanical and other factors. [↑](#endnote-ref-10)
11. For a rather different and interesting account of the implications of this "conceptual" fact that we must say that any intentional action results from a want or desire, see G. F. Schueler, Desire: Its Role in Practical Reason and the Explanation of Action (Cambridge, Mass.: MIT Press, 1995. [↑](#endnote-ref-11)
12. And notice that this is a very general problem for any theory that uses the notion of desire. This includes mixed views which use the notion of desire to explain some but not all of the motivational phenomena. [↑](#endnote-ref-12)
13. Smith, "The Humean Theory of Motivation," p. 54. [↑](#endnote-ref-13)
14. Huw Price makes this point in "Defending Desire as Belief" (Mind 98 [1989]: 119-127) in relation specifically to Smith's particular spelling out of the direction of fit argument. But, as I shall show below, the general point can be made with regard to other versions of the direction of fit argument as well. [↑](#endnote-ref-14)
15. Lloyd Humberstone, "Direction of Fit" (Mind 101 [1992]: 59-83). [↑](#endnote-ref-15)
16. To make this precise, we would have to qualify these preferences so that they were prima facie preferences. But I leave this detail aside here and in what follows. [↑](#endnote-ref-16)
17. The literature seems to be divided on whether the notion of direction of fit is essentially confused. See, for example, G. F. Schueler, "Pro-Attitudes and Direction of Fit" Mind 100 (1991): 277-81 for a critique and Lloyd Humberstone, "Direction of Fit," Mind 101 (1992): 59-83 for a defense. [↑](#endnote-ref-17)
18. Similarly, we might have metaphysical scruples about positing "normative facts" that would be the objects of ought-beliefs. Thus to avoid having an ethics that either countenances bare normative facts or which claims that all moral beliefs are false (since there are no bare normative facts and since, therefore, no state of affairs has any normative properties), it may seem as though we must not use the cognitivist vocabulary. I thank Don Hubin for pointing this out to me. [↑](#endnote-ref-18)
19. I am very grateful to Walter Edelberg and Al Mele for extensive comments on drafts of this paper and very useful discussions about the issues it raises. Don Hubin and an anonymous reviewer for *Philosophical Studies* also provided useful comments on the penultimate draft. Several of my fellow graduate students at the University of Illinois at Chicago also provided useful input on very early versions of this material; Samantha Brennan and Charlie Huenemann were particularly helpful. An ancestor of this paper was presented to the Graduate Student Philosophy Conference at Syracuse University, March, 1993. I thank Neil Manson (who was the commentor) and other participants for useful input. Another version was presented at the Pacific Division Meetings of the American Philosophical Association in Los Angeles, March 31, 1994. I thank Garret Deckel (who was the commentor) and other participants for valuable feedback. I am grateful to the APA for presenting me with a Graduate Student Essay Award which enabled me to travel to Los Angeles to deliver the paper. Some work on this paper was done while I was on a Dean's Scholar Fellowship at the University of Illinois at Chicago, and I am grateful for that support. [↑](#endnote-ref-19)