

COGNITIVIST VS NON-COGNITIVIST EXPLANATIONS OF THE BELIEF-
LIKE AND DESIRE-LIKE FEATURES OF EVALUATIVE JUDGEMENT*

Michael Smith

When an agent judges her performance of some action to be desirable she thereby conveys her assessment of the importance of acting in the way in question, an assessment in which she might have more or less confidence and for which she might have more or less justification (Brink 1989). To this extent an agent's evaluative judgements seem to express her beliefs. At the same time, an agent's assessment of the importance of acting in a certain way seems to correlate in a fairly robust way with her motivational potential, at least in so far as she is rational (Hare 1952). No desire to (say) do whatever she happens to value doing seems needed to lead a rational subject who judges it desirable to act in a certain way to be motivated to act in that way, and, indeed, to be more motivated to do that which judges more desirable and less motivated to do what she judges less desirable. To this extent, an agent's evaluative judgements seem to express her desires.

The fact that evaluative judgements have both belief-like and desire-like features leads to a common assumption, an assumption that seems to me to structure much of the debate in contemporary meta-ethics (see also Smith 1994, Ch.1). The assumption is that whereas cognitivism — the view evaluative judgements express beliefs — has an easy time accommodating the belief-like features of evaluative judgement, it faces an insuperable difficulty in accounting for the desire-like features. Non-cognitivism — the view that evaluative judgements express some non-belief state; or, more precisely, that version of non-cognitivism according to which evaluative judgements express desires — is supposed to have the opposite problem. While it has an easy time explaining the connection between evaluative judgement and motivation — desire is, after all, a motivating state — it is supposed to face an insuperable difficulty in accounting for the potential for justification and rational defence that such motivational antecedents must therefore have.

* This paper draws on material that appears in Smith 2002a and Smith 2002b. I would like to thank Simon Blackburn, Michael Bratman, Mark Greenberg, Lloyd Humberstone, Frank Jackson, James Lenman, Philip Pettit, Michael Ridge, and Jack Smart for their useful comments.

My suspicion is that this assumption is simply mistaken. There is at least one version of cognitivism that can accommodate both the potential that evaluative beliefs have for justification and rational defence and the connection between evaluative beliefs and motivation. Moreover no version of non-cognitivism — or, anyway, no version with which I am familiar — is able to accommodate either of these features. This rather sweeping negative claim about non-cognitivism might be wrong, of course. Perhaps there is some unknown form of non-cognitivism waiting in the wings that can explain either the belief-like features, or the desire-like features, or perhaps even both. But if there is then I suspect that that will be because this more plausible form of non-cognitivism differs in some fundamental way from the versions that are currently on offer. I will be happy enough if the effect of this paper is to force non-cognitivists to come up with a more plausible formulation.

The paper is in five main sections. In the first I describe a version of cognitivism that is capable of explaining both the belief-like and the desire-like features of evaluative judgement. In the second I describe what seems to me to be one of the best-known and most sophisticated versions of non-cognitivism — the version outlined by Simon Blackburn in his Ruling Passions (1998) — and I explain why, if Blackburn's theory is to accommodate the belief-like features of evaluative judgement, it must collapse into the preferred version of cognitivism. In the third section I describe some more complex belief-like features of evaluative judgement and the way in which these impact on the explanation of action. In the fourth section I explain why the version of cognitivism described in the first section can handle these complexities in a natural and compelling way. In the fifth section I explain why non-cognitivism seems to be unable to accommodate these complexities.

1. The preferred form of cognitivism and why it can explain both the belief-like and the desire-like features of evaluative judgement

Cognitivists claim that when an agent judges some action to be desirable, she thereby expresses a belief she has about the action in question. But what belief does she thereby express? Can we say something more illuminating about its content?

It seems to me helpful, in this connection, to start from the more or less common sense assumption that for an agent to believe that her acting in a certain way in certain

circumstances is desirable is for her to believe that her so acting in those circumstances is advisable: a matter of her believing that she would advise herself to perform that act in those circumstances if she were herself in circumstances in which she was best placed to give herself advice (Smith 1995). This is the appropriate place to start because the fact that one would give oneself such advice in such circumstances clearly has normative force. Two questions then come naturally to mind. First, what are these circumstances in which agents are best placed to give themselves advice? Second, what fixes the content of the advice that the agents in those circumstances would give to themselves?

The answer to the first question is, I suggest, that agents are best placed to give themselves advice when their psychologies have been purged of all cognitive limitations and rational failings, and the answer to the second is that the content of such advice is then fixed by the contents of the desires that such agents, with their psychologies suitably purged, would have about what they are to do in the circumstances of action about which they are seeking advice. In other words, when I judge my performance of a certain action to be desirable that amounts to my believing that my performance of that act is advisable, where that, in turn, amounts to my believing that I would want myself so to act if I had a desire set that was purged of all cognitive limitations and rational failings.

If something like this is along the right lines then all we need in order to get a full-blown analysis of desirability is to give an account of the conditions that need to be met by a desire set which is devoid of cognitive limitations and rational failings. My suggestion in this regard, developing an idea of Bernard Williams's (1980), is that for a desire set to be devoid of cognitive limitations and rational failings is for it to be one which meets certain descriptive-cum-normative conditions: specifically, it is for one's desire set to be maximally informed and coherent and unified (Smith 1994). If we call the possible world in which the agent has the desires that she actually has in the circumstances of action she faces the 'evaluated' world, and the possible world in which she has that set of desires that is maximally informed and coherent and unified the 'evaluating' world, then, the suggestion is, what it is desirable for her to do in the evaluated world is fixed not by what, in the evaluated world she wants herself to do in the evaluated world, and not by what, in the evaluating world, she wants herself to do

in the evaluating world, but rather by what, in the evaluating world, she wants herself to do in the evaluated world. This, accordingly, is the property that an agent must believe her act to have when she values the performance of that act.

Once this is agreed, it seems to me that there is no difficulty at all in seeing, at least in broad terms, why evaluative judgement has both belief-like and desire-like features. Since judging it desirable to act in a certain way is a matter of having certain beliefs about what it is desirable for the agent to do, where desirability is in turn a matter of what the agent would want herself to do if she had a set of desires that was maximally informed and coherent and unified, it should come as no surprise at all to learn that her evaluative judgements convey her assessment of the importance of acting in the way in question, where this assessment is one in which she might have more or less confidence and for which she might have more or less justification. An agent's evaluative beliefs are, after all, just a species of belief, so the reason that these features are features of evaluative beliefs is that they are features of beliefs quite generally.

But nor should it come as any surprise to learn that an agent who judges it desirable to act in a certain way has a belief that is capable of both causing and rationalizing certain desires without the aid of any further desire, a desire such as, for example, the desire to do what she values (Smith 2001a). In order to see that this is so, imagine a case in which, on reflection, you come to believe that (say) you would desire that you experience pleasure in the circumstances that you presently face if you had a maximally informed and coherent and unified set of desires, but imagine further that you don't in fact have any desire at all to experience pleasure in these circumstances, or perhaps that you are averse to doing so. Now consider the pair of psychological states comprising, on the one hand, your belief that you would desire that you experience pleasure in the circumstances that you presently face if you had a maximally informed and coherent and unified set of desires, and, on the other, the desire that you experience pleasure in these circumstances, and compare this pair of psychological states with the pair comprising that belief together with aversion to experiencing pleasure in these circumstances, and the pair comprising that belief together with indifference. Which of these pairs of psychological states is more coherent?

The answer would seem to be plain enough. Indifference and aversion are attitudes that you believe you wouldn't have if you had a more coherent desire set. To this extent they must therefore fail to cohere with that belief. (Think of an analogy. The belief that you would believe that *p* if you had a more coherent belief set coheres better with believing that *p* than it does with disbelieving *p*, or with having no view about whether or not *p*.) There is therefore more than a passing family resemblance between the relation that holds between the first pair of psychological states and more familiar examples of coherence relations that hold between psychological states. Coherence would thus seem to be on the side of the pair comprising both the belief that you would desire that you experience pleasure in the circumstances that you presently face if you had a maximally informed and coherent and unified desire set and the desire that you experience pleasure. But, if this is right, then it follows immediately that in so far as you are rational, in the relatively mundane sense of having and exercising a capacity to have the psychological states that coherence demands of you, then you will end up having a desire that matches your belief about what you would want if you had a maximally informed and coherent and unified desire set. In other words, in the particular case under discussion, you will end up losing your aversion or indifference to experiencing pleasure, and acquiring a desire to experience pleasure instead.

The belief that you would desire that things be a certain way if you had a set of desires that was maximally informed and coherent and unified would thus seem able to cause you to acquire a corresponding desire when it operates in conjunction with the capacity to have coherent psychological states. Moreover, because acquiring that desire makes for a more coherent pairing of psychological states, it would seem to follow that the desire thus caused is rationalized as well. Finally, note that no causal role at all needs to be played by any desire in the explanation of the acquisition of this desire. All that is required is the exercise of the capacity to have coherent psychological states, a capacity whose exercise is ubiquitous across both the cognitive and the non-cognitive realms.

Let me sum up the argument in this section. A certain version of cognitivism, one according to which evaluative judgements are beliefs about what we would want ourselves to do if we had a maximally informed and coherent and unified desire set,

would seem capable of explaining both the belief-like and the desire-like features of evaluative judgements. The common assumption in meta-ethics that cognitivist theories are incapable of explaining the desire-like features of evaluative judgement thus seems to me to be mistaken.

2. Problems for the non-cognitivist's explanation of the belief-like features of evaluative judgement

It is important to note, from the outset, the difficulty involved in supposing that evaluative judgements express desires or aversions, rather than beliefs. As we have seen, the psychological states that we express when we make evaluative judgements are remarkably belief-like. I take this to mean that they have many of the functional features of belief. The difficulty, to anticipate, is to conceive of desires or aversions that have exactly these features, given that the main difference between beliefs, on the one hand, and desires and aversions, on the other, is that they differ in just these functional respects. Let me explain.

The function of a belief is to represent things as being a certain way. Beliefs manage to do this, in part, by coming prepackaged with links to other beliefs and perceptions that serve as sources of epistemic support. In the absence of these sources of epistemic support it is the role of beliefs simply to disappear. To believe something at all is thus to believe a whole host of things which, together, are supposed to provide some sort of justification for what is believed. Desires, by contrast, are the exact opposite of beliefs in this respect. The function of a desire is not to represent things as being a certain way, but rather (very roughly) to represent things as being the way they are to be. Desires thus do not come prepackaged with links to other desires which provide them with (some analogue of) epistemic support. Instead they come prepackaged with the potential to link up with beliefs about means so as to produce action, and in the absence of which they remain (more or less) dormant.

Now consider the psychological states that we express when we make evaluative judgements. These psychological states are plainly functionally like beliefs in that they come prepackaged with links to other psychological states that provide them with (some analogue of) epistemic support. For example, to accept the desirability of keeping promises is to be disposed to accept a whole host of claims which, as it seems

to us, together provide this claim with some sort of justification, and in the absence of which acceptance of the desirability of keeping promises would simply disappear. This is the reason that the procedure of reflective equilibrium Rawls describes sounds so platitudinous (Rawls 1951). The reflective equilibrium procedure simply takes to the limit the commonplace procedure whereby we test the various particular evaluative judgements we are disposed to make against the host of other evaluative (and non-evaluative) claims we accept and from which the various particular evaluative judgements we are disposed to make gain their (analogue of) epistemic support.

It should now be plain why there is a special difficulty involved in finding desires or aversions with belief-like features. For the difference between beliefs and desires and aversions is, *inter alia*, that beliefs do, whereas desires and aversions do not, come prepackaged with links to other desires and aversions which provide them with (some analogue of) epistemic support. This is why the unwilling addict can rightly claim that it is in no respect desirable for him to take drugs, notwithstanding the fact that he desires to take them (compare Watson 1975). The unwilling addict's desire for drugs exists entirely independently of any links to desires or aversions that provide it with some analogue of epistemic support. It is simply a brute drive caused in him by his consumption of drugs. When a non-cognitivist tries to spell out the nature of the special desires and aversions that we express when we make evaluative judgements, then, he must therefore take care to stipulate the presence of such links.

Notwithstanding the fact that desires and aversions do not come prepackaged with links to other desires and aversions that provide them with some analogue of epistemic support, it is thus crucial to the truth of non-cognitivism that such links do exist. But nor should we be skeptical about this. We in effect saw this when we spelled out the preferred form of cognitivism above. Certain links between desires and aversions do indeed provide them with an analogue of epistemic support (Williams 1980; Smith 1994, 155-161).

For example, desires and aversions, like beliefs, are sensitive to information. The fact that a particular desire or aversion is uninformed and would disappear upon the impact of information — imagine someone who has a preference for drinking red

wine over white, but who is ignorant of the taste of both, and yet would hate red wine if he tasted it and enjoy white — is ordinarily taken to be grounds for rational criticism of that desire. Desires and aversions, like beliefs, also fit together in more and less coherent ways. The fact that a particular desire or aversion contributes incoherence to the overall set of desires and aversions of which it is a member is thus also ordinarily taken to be grounds for rational criticism. Finally, desires and aversions fit together with other desires and aversions in more or less unified ways, much as beliefs fit together with other beliefs in more or less unified ways. The fact that a desire or aversion contributes disunity to the overall set of desires and aversions of which it is a member is thus also ordinarily taken to be grounds for rational criticism.

What non-cognitivists must stipulate, then, when they try to spell out the special nature of the desires and aversions that we express when we make evaluative judgements, is that these desires and aversions are parts of sets of such special desires and aversions that are, in turn, sufficiently informed and coherent and unified, where to say that a set of desires and aversions is 'sufficiently' informed and coherent and unified is simply to say that the links among the desires and aversions in that set are similar in number and quality to the minimum number and quality of epistemic links among beliefs that we are prepared to tolerate for the ascription of a belief. To repeat, they must stipulate such links on pain of failing to capture the functional role of the psychological states that we express when we make evaluative judgements.

Now, as I understand it, one of the world's best-known and most sophisticated non-cognitivists, Simon Blackburn, undertakes the task of spelling out the nature of the special desires and aversions that we express when we make evaluative judgements in his new book Ruling Passions (Blackburn 1998). Let's therefore consider his proposal in the light of these difficulties.

We should think in terms of a staircase of practical and emotional ascent. At the bottom are simple preferences, likes, and dislikes. More insistent is a basic hostility to some kind of action or character or situation: a primitive aversion to it, or a disposition to be disgusted by it, or to hold it in contempt, or to be angered by it, or to avoid it. We can then ascend to reactions to such reactions.

Suppose you become angry at someone's behavior. I may become angry at you for being angry, and I may express this by saying it is none of your business. Perhaps it was a private matter. At any rate, it is not a moral issue. Suppose, on the other hand, I share your anger or feel 'at one' with you for so reacting. It may stop there, but I may also feel strongly disposed to encourage others to share the same anger. By then I am clearly treating the matter as one of public concern, something like a moral issue. I have come to regard the sentiment as legitimate. (Blackburn 1998, 9)

As I read this passage, the last sentence is supposed to be the conclusion of an argument whose premises are spelled out earlier on in the passage. Blackburn continues:

Going up another step, the sentiment may even become compulsory in my eyes, meaning that I become prepared to express hostility to those who do not themselves share it. Going up another level, I may also think that this hostility is compulsory, and be prepared to come into conflict with those who, while themselves concerned at what was done, tolerate those who do not care about it. I shall be regarding dissent as beyond the pale, unthinkable. This should all be seen as an ascending staircase, a spiral of emotional identifications and demands. The staircase gives us a scale between pure preference, on the one hand, and attitudes with all the flavor of ethical commitment, on the other. (Blackburn 1998, 9)

And again, as he puts it later:

To sum up, then: to hold a value is to have a relatively fixed attitude to some aspect of things, an attitude with which one identifies in the sense of being set to resist change, or set to feel pain when concerns are not met. That fixed attitude typically issues in many dispositions, at various places on the staircase of emotional ascent I described. (Blackburn 1998, 68)

Blackburn thus clearly thinks that he has made a case for the idea that someone who has the higher order attitudes and dispositions he describes is in a state of mind that is best described as a normative commitment: having the higher-order attitudes entails

having the normative commitment. These are therefore the special attitudes he thinks get expressed when we make evaluative judgements. As is perhaps already clear, however, Blackburn's official story is not entirely convincing.

We can, after all, readily imagine someone who (say) has a desire that people keep their promises, and who shares many other people's anger at those who fail to keep their promises, and who feels disposed to encourage others to share that same anger too, and who feels disposed to be angry at those who don't share that anger, and yet who doesn't regard any of these sentiments as being in the least legitimate. We need simply to imagine someone who, in addition, regards all of his various attitudes towards promising in much the same way as the unwilling addict regards his addiction. He might think, for example, that these attitudes were all simply caused in him by social forces, in much the same way as the ingestion of drugs caused the unwilling addict's desire to take drugs in him, and that no reasons can therefore be given in support of acting on the basis of these attitudes, much as the addict thinks that no reasons can be given for his acting on his desire to take drugs.

More generally, then, it would seem that nothing about the mere location of attitudes on the staircase of emotional and practical ascent Blackburn describes guarantees that the agent who has those attitudes takes them to be justified. Nothing about the mere location of attitudes on such a staircase suggests that they are parts of sets of desires and aversions that are sufficiently informed and coherent and unified. To this extent Blackburn's account fails to capture the functional role of the psychological states that we express when we make evaluative judgements.

How might Blackburn reply to this objection? He might reply that he has indeed told us a story about the justification of our normative commitments, because such a story has nothing at all to do with whether or not the desires and aversions that constitute our normative commitments are parts of sets of desires and aversions that are sufficiently informed and coherent and unified. Instead, he might say, it has everything to do with whether or not we are normatively committed to having such normative commitments. We not only have these desires, but desire to have them, and desire to desire to have them, and so on. The image of a staircase of emotional

and practical ascent is thus, he might insist, on exactly the right track (Blackburn 2002, 176).

But this reply plainly rests on a pun on the word 'justification'. It conflates two quite separate questions we can ask about the psychological states that constitute our normative commitments. One question we can ask is whether we judge it desirable to have such psychological states. Another is whether we take such psychological states to enjoy (some analogue of) epistemic support. In these terms, the crucial point is that though it would seem to be constitutive of the psychological states that constitute our normative commitments that we take them to have the belief-like feature of enjoying (some analogue of) epistemic support, it is not constitutive of such psychological states that we ascribe them any particular normative status. Yet this is what the imagined reply presupposes. We may think a normative commitment epistemically justified and yet, unfortunately, an undesirable normative commitment to have. Indeed, those consequentialists who are convinced that the theory is self-effacing think that something like this is in fact true of a commitment to the truth of consequentialism (Parfit 1984, 40-43; Pettit and Smith 2000; Smith 2001b).

An alternative reply is suggested by the earlier discussion of the preferred form of cognitivism. It might be thought that Blackburn could simply stipulate that the syndrome of attitudes that he says gets expressed in evaluative judgements are parts of such sets. I will consider this reply below. But, even conceding for a moment that this reply succeeds, the very fact that he needs to make the reply must surely be an embarrassment. For it amounts to an admission that what makes a desire justified — in the sense of 'justification' that has its analogue in the notion of epistemic support — and hence immune to rational criticism, is its being part of a set of desires that is informed and coherent and unified. This is tantamount to admitting that at least one evaluative judgement — the claim that a desire is justified, or immune from criticism — is the expression of a belief, rather than a desire. This is cognitivism, not non-cognitivism.

Let's now consider the view about the nature of the special desires and aversions we express when we make evaluative judgements to which Blackburn would be committed if he were to make the alternative reply. When we make such claims

Blackburn would have to say that we thereby express certain desires or aversions that are located in his preferred place on the staircase of emotional and practical ascent, but he would then have to add that these desires and aversions exist alongside other similarly located desires and aversions which are such that, together, they make up a sufficiently informed and coherent and unified set. Would this solve the problem? I do not think so.

The reason is that we could evidently believe that we have such desires and aversions, and hence believe that they are justified or immune to criticism relative to the sufficiently informed and coherent and unified set of which they are members, and yet also believe that we would not have such desires and aversions if we had a set of desires and aversions that was maximally informed and coherent and unified. (A coherent sub-set of desires and aversions might fail to cohere with the larger set of which it is a sub-set.) In that case I take it that we would not believe that the original desires and aversions were justified or immune from criticism. We would rather believe the opposite.

The upshot would therefore seem to be that since, when we claim that it is desirable or undesirable to act in a certain way, we must suppose that we have a justification for making this claim, and since, by the non-cognitivist's own lights, we could not suppose that we have such a justification if we believed that the desires or aversions that we thereby express would be no part of a maximally informed and coherent and unified set of desires and aversions, so, when we make such a claim, we must at least implicitly believe that the desires or aversions we thereby express would be part of a maximally informed and coherent and unified set of desires and aversions. Only so does it appear that we could take ourselves to have the sort of justification we require.

However, if this is right, then it follows that Blackburn must further revise his account of the nature of the special desires and aversions that we express when we make evaluative judgements. He must suppose not just that we express certain desires or aversions that are located in his preferred place on the staircase of emotional and practical ascent he describes, and not just that these desires and aversions exist alongside other similarly located desires and aversions which, together, make up a sufficiently informed and coherent and unified set. He must suppose that, in addition,

these desires and aversions are possessed in the presence of at least an implicit belief that such desires and aversions would be possessed if we had a set of desires and aversions that was maximally informed and coherent and unified.

However, if Blackburn were to revise his account of what makes desires and aversions special in this way then it is unclear whether the desires and aversions he says we have when we have normative commitments any longer play an essential role in the explanation of the belief-like functions of our normative commitments. These functions could, after all, now be explained entirely by the presence of the belief whose presence must be postulated. And nor are the desires and aversions Blackburn says we have when we have normative commitments required to explain the desire-like features of our normative commitments either. For the reasons given when we spelled out the preferred form of cognitivism, this too can be explained by the belief whose presence we must now postulate, together with the fact that agents have a general capacity to acquire the beliefs and desires that are mandated of them by considerations of coherence. But if the desires and aversions Blackburn postulates play no role in an explanation of either the belief-like or the desire-like functions of the psychological states we express when make evaluative judgements, then their postulation is completely idle. Blackburn's non-cognitivism collapses into cognitivism.

Let me sum up the argument in this section. We saw that the fact that evaluative judgements have belief-like features puts constraints on the kind of state could be expressed when we make an evaluative judgement. We considered the account of the nature of such states provided by one of the world's best-known and most sophisticated non-cognitivists, namely, Simon Blackburn. We saw that his account has a fatal flaw, and, in order to repair that flaw, he would in effect have to abandon his non-cognitivism in favour of the preferred form of cognitivism.

In the next section I will spell out three further belief-like features of evaluative judgements and explore the ways in which these features interact with the desire-like features of such judgements. To anticipate, my argument will be that the preferred form of cognitivism can, and that non-cognitivism cannot, accommodate the complex ways in which these features suggest evaluative judgements connect with motivation.

For this reason it will turn out that cognitivism can, and that non-cognitivism cannot, explain the desire-like features of evaluative judgement. The upshot will therefore be that non-cognitivism can explain neither the belief-like nor the desire like features of evaluative judgements.

3. How the belief-like and the desire-like features of evaluative judgement interact

Commonsense tells us that evaluative judgements have three quite further belief-like features, features in addition to that already described. Two of these are features that such judgements share with beliefs quite generally. The third is a feature that is distinctive of evaluative judgements in particular. Let me describe each of these three belief-like features more fully.

The first feature of an evaluative judgement is the level of confidence a subject has that things are evaluatively as she judges them to be. For example, a subject might think it very unlikely that (say) being honest is desirable in itself, but much more likely that being knowledgeable is desirable in itself, and even more likely still that experiencing pleasure is desirable in itself. This is, as I said, a feature that evaluative judgements share with beliefs quite generally. To take a non-evaluative example, a subject might think it very unlikely that the sun will explode tomorrow, but much more likely that it will rain tomorrow, and even more likely still that there will be a football match tomorrow.

The fact that subjects can have different levels of confidence in both evaluative and non-evaluative judgements alike is important, because it reminds us that the level of confidence that a subject has in certain of her evaluative judgements is a product of both a level of confidence she has in another evaluative judgement and the level of confidence she has in a non-evaluative judgement. For example, the level of confidence a subject has that her riding on a roller coaster is desirable might be a product of two levels of confidence: first, her level of confidence in the non-evaluative judgement that her riding on a roller coaster would lead her to experience pleasure, and, second, her level of confidence in the evaluative judgement that experiencing pleasure is desirable. Importantly, however, the levels of confidence that subjects have in certain of their evaluative judgements are not in this way the products of the confidence they have in other evaluative judgements and the

confidence they have in some non-evaluative judgement. The level of confidence a subject has in the judgement that pleasure is desirable in itself is, we may assume, like this. It is this kind of evaluative judgement, a judgement of intrinsic or fundamental or non-derived value, which will be the focus of discussion in the remainder of this paper.

How do we measure the different levels of confidence subjects have in their judgements of intrinsic value? Differences in subjects' levels of confidence is the sort of thing that gets revealed in how much they would be willing to bet on one outcome as opposed to another under circumstances of forced choice. It would perhaps be difficult to construct an appropriate betting situation that would reveal the different levels of confidence that subjects have in the claims they make about what is of intrinsic value, but the basic idea should in principle be clear enough. Facts about subjects' different levels of confidence are thus, as we might put it, synchronic facts about their evaluative judgements. In referring back to this feature of evaluative judgements later, I will call this their 'Certitude'.

The second feature of evaluative judgement, another that such judgements share with beliefs in general, concerns how stable a subject's confidence that things are evaluatively the way that she judges them to be is under the impact of incoming information and reflection. For example, though a subject might be equally confident that (say) experiencing pleasure is desirable in itself and being autonomous is desirable in itself, her confidence in the former might be very stable under the impact of incoming information and reflection, whereas her confidence in the latter is very unstable.

As I said, this too is a feature that evaluative judgements share with beliefs in general. Again, to take a non-evaluative example, though a subject might have the very same level of confidence — or, as we might say, the very same degree of belief — in the proposition that the Sydney Swans is the most talented Australian Rules football team and the proposition that her son is a responsible supermarket employee, her degree of belief in the proposition that her son is a responsible supermarket employee might be very stable under the impact of incoming information and reflection, whereas her degree of belief in the proposition that the Sydney Swans is the most talented

Australian Rules football team is much more unstable. Though nothing much in the way of incoming information and reflection would change her confidence in the proposition that her son is a responsible supermarket employee all that much — perhaps the evidence is already in — all sorts of incoming information would radically change her confidence in the proposition that the Sydney Swans is the most talented Australian Rules football team.

Whereas facts about subjects' levels of confidence are fixed synchronically by how much they would bet on the propositions they accept under circumstances of forced choice, facts about how stable the levels of confidence in the propositions subjects accept are under the impact of incoming information and reflection are plainly fixed diachronically. These facts are fixed by changes in how much subjects would be willing to bet on one outcome as opposed to another over time. In order to be able to refer back to this feature of evaluative judgement later, let's call this their 'Robustness'.

The third feature of evaluative judgements on which I wish to focus is a feature of such judgements in particular. It concerns, specifically, what such judgements are about. When a subject judges that (say) both experiencing pleasure and being autonomous are desirable in themselves, it is always relevant to ask how desirable she judges each of these features to be: which she judges to be more desirable. The best way to fix on this feature of evaluative judgements is to imagine the perspective of people who are omniscient, for we are then able to abstract away from any differences due to Certitude. As between two outcomes, in one of which they are autonomous to a certain extent, and in the other of which they experience a certain amount of pleasure, how autonomous would people who are omniscient have to imagine themselves being, as compared with experiencing how much in the way pleasure, in order for them to be indifferent? Intuitively, the idea is, the less autonomous they would have to imagine themselves being, and the more they would have to imagine themselves experiencing in the way of pleasure, in order to be indifferent between being autonomous to that extent as opposed to experiencing that amount of pleasure, the more desirable they would thereby be taking autonomy to be and the less desirable they would thereby be taking experiencing pleasure to be. In order to refer back to this feature later, let's call this feature of evaluative judgements their 'Importance'.

To sum up what we have said so far: we have distinguished three different features commonsense tells us evaluative judgements have, in so far as they have belief-like features. These are Certitude, Robustness, and Importance. I suggested at the beginning of this section that these three features are crucially important when it comes to understanding the way in which an agent's values explain her actions. We are now in a position to see why that is so.

Our initial statement of the desire-like feature of an evaluative judgement was that, in so far as agents are rational, we expect them to be motivated to do what they judge it desirable to do. However, as is perhaps now evident, this initial statement is far too crude. It is far too crude because it abstracts away from Certitude, Robustness and Importance as features of evaluative judgements, whereas commonsense tells us that the motivations of a rational agent will plainly be crucially dependent on these features. In order to be more accurate we must therefore state the desire-like features of evaluative judgements in the following rather more complicated terms.

To begin with, in so far as they are rational, agents will be more strongly motivated to do that which they judge it more desirable to do, as between options about which their confidence levels are the same. In other words, in the terms just introduced, at a time, the strengths of a rational agent's different motivations will reflect such differences as might exist in relative Importance, abstracting away from differences in Certitude. For example, if a rational agent is equally confident that experiencing a certain amount pleasure is desirable in itself and that being autonomous to a certain degree is desirable in itself, but judges that being autonomous to that degree is more desirable than experiencing pleasure to that extent, then she will be more strongly motivated to be autonomous to that degree than to experience pleasure to that extent.

Second, in so far as they are rational, agents are more strongly motivated to do that about which they are more confident, as between options which they judge to be equally desirable. In other words, in the terms just introduced, at a time, we expect that the strength of a rational agent's different motivations will covary with such differences as might exist in Certitude, abstracting away from differences in Importance. For example, if a rational agent judges that experiencing pleasure to a certain extent and being knowledgeable to a certain extent are equally desirable, but

she is much more confident that experiencing pleasure to that extent is desirable than she is that being knowledgeable to that extent is desirable, then she will be more strongly motivated to experience the pleasure than she is to be that knowledgeable.

Finally, in so far as they are rational, the stability of agents' motivations over time covaries with the stability of their evaluative judgements. In other words, in the terms just introduced, over time, we expect the stability of a rational agent's motivations to track Robustness, abstracting away from such differences as might exist in Importance and Certitude. For example, if a rational agent judges it desirable to a certain degree to experience pleasure, and if her confidence level is stable under the impact of information and reflection, then the strength of her motivation to experience pleasure will be stable under the impact of information and reflection. But if a rational agent judges it desirable to a certain degree to be autonomous, and if her confidence level waxes and wanes under the impact of information and reflection, then the strength of her motivation to be autonomous will wax and wane under the impact of information and reflection too.

We now have before us our more complicated formulation of the desire-like features of evaluative judgement. To repeat, this more complicated formulation of the desire-like features of evaluative judgement is mandated by the three further belief-like features that evaluative judgement are supposed to have, according to commonsense. It thus follows that, to the extent that a theory about the nature of evaluative judgement is unable to give a satisfactory account of the three features, and the roles that they play in the explanation of action, that theory fails to accord with commonsense. The question we must ask ourselves is thus whether cognitivists and non-cognitivists are equally able to make room for these three features of evaluative judgement and whether they are also equally able to accommodate the crucial roles played by these three features when it comes to understanding the way in which an agent's values explain her actions.

In the next section I consider how the preferred form of cognitivism described earlier makes room for the three further belief-like features of evaluative judgements just described and whether it can accommodate the ways in which these features impact on the explanation of action. In the section after that I say a little about how non-

cognitivists might attempt to make room for the three further belief-like features and accommodate their different roles.

4. How the preferred form of cognitivism accounts for the three further belief-like features of evaluative judgement and the different roles they play in the explanation of action

In an earlier section I outlined a preferred form of cognitivism. This allowed us to see, in broad outline, why, in so far as we are rational, our evaluative beliefs give rise to corresponding motivations. But once we have seen that story in broad outline it seems to me that we can also see why the more complex belief-like features that our evaluative beliefs possess — Certitude, Robustness and Importance — impact on our motivations in the way that they do.

Consider Importance first. Suppose an agent believes that if she had a maximally informed and coherent and unified desire set then she would desire that she experiences a certain amount of pleasure in the circumstances that she presently finds herself in, but she also believes that she would have a stronger desire that she be autonomous to a certain degree in these circumstances. In other words, more intuitively, suppose that she believes that being autonomous to that degree is more important than experiencing that amount of pleasure. What does coherence require of her?

Coherence plainly demands of her not just that she has desires both to experience that amount of pleasure and to be autonomous to that degree in the circumstances in which she presently finds herself, but that the relative strength of her desires matches the relative strength of the desires she believes she would have: demands of her, in other words, that her desire to be autonomous to that degree be stronger than her desire to experience that amount of pleasure. The argument for this is simply a more sophisticated version of the argument given earlier.

Consider the quadruple of psychological states that comprises the agent's belief that she would have a desire of a certain strength that she experiences the relevant amount of pleasure in the circumstances in which she presently finds herself if she had a maximally informed and coherent and unified desire set, and the belief that she would

have a somewhat stronger desire that she be autonomous to the relevant degree in these circumstances, together with a desire that she experiences pleasure in these circumstances and a correspondingly stronger desire that she be autonomous. This quadruple of psychological states would seem to exhibit more in the way of coherence than (say) the quadruple of psychological states that comprises those beliefs together with a stronger desire that she experiences pleasure and a weaker desire that she be autonomous.

Now consider Certitude. Imagine the agent just described with the beliefs just described, but let's suppose further while she has a certain high degree of confidence that she would desire that she experience pleasure in the circumstances in which she presently finds herself if she had a maximally informed and coherent and unified desire set, her confidence that she would have a somewhat stronger desire that she be autonomous in the circumstances that she presently finds herself is somewhat lower. In other words, put more intuitively, though she believes that being autonomous is more important than experiencing pleasure, she is far more confident of her assessment of the value of experiencing pleasure than she is of her assessment of the value of being autonomous. What does coherence demand of her in this case?

Again, it seems plain enough that we could construct a similar comparison of quadruples of psychological states to show that coherence demands of her that the relative strengths of her desires to be autonomous and experience pleasure matches the relative strengths of the desires she believes that she would have that she be autonomous and experience pleasure, but this time, as discounted by her different levels of confidence. Indeed, we can even imagine the level of confidence that she would desire that she be autonomous being so low, as compared with her level of confidence that she would desire that she experiences pleasure, that, notwithstanding the fact that she believes that her desire that she be autonomous would be stronger than her desire that she experience pleasure, coherence may even demand of her that her actual desire that she experience pleasure be stronger than her desire that she be autonomous.

Consider, finally, Robustness. Imagine the agent just described, with the confidence levels just described, but let's suppose further that the higher level of confidence she

has that she would desire that she experience pleasure in the circumstances in which she presently finds herself if she had a maximally informed and coherent and unified desire set diminishes over time under the impact of information and reflection, and that her lower level of confidence that she would have a somewhat stronger desire that she be autonomous in the circumstances that she presently finds herself increases over time under the impact of information and reflection. What does coherence demand of her over time in this case? The answer is, again, plain. Coherence demands of her over time that her actual desires that she experience pleasure and be autonomous shift in their relative strengths so that they reflect, at each moment, the appropriate mix of believed strength and level of confidence.

We have thus seen that at least one cognitivist account of evaluative judgement, the account that I myself prefer, can not only make room for the three features of such judgements that we described earlier, and accommodate the different roles that these features play in the explanation of action, but that it can also do so in a straightforward and intuitive way. This particular version of cognitivism thus squares extremely well with commonsense. The question we must now ask is whether the same can be said for a non-cognitivist account of evaluative judgement.

5. How non-cognitivism accounts for the three further belief-like features of evaluative judgement and the different roles they play in the explanation of action

How might non-cognitivists attempt to make room for Certitude, Importance, and Robustness as features of evaluative judgement? The answer to this question is crucially important. For without an answer to this question they are plainly unable to accommodate the different roles that these features play in the explanation of action.

Remember, non-cognitivists hold that a subject's evaluative judgements are expressions not of her beliefs about the ways that things in evaluative respects, but are rather expressions of her desires that things be a certain way in non-evaluative respects. The judgement that experiencing pleasure is desirable in itself, for example, is claimed by non-cognitivists not to be the expression of a belief about an evaluative property, desirability, that is possessed by the experience pleasure, but rather to be the expression of desire to experience pleasure, or a desire to desire to experience pleasure, or some other, similar, non-belief state. It thus follows that non-cognitivism

labours under a difficulty in explaining the three features of evaluative judgement, a difficulty not faced by cognitivists.

Cognitivists can explain at least one of the three features of evaluative judgements, namely Importance, in terms of a feature of the world: think again of the explanation given above according to which Importance is a matter of desirability, conceived of as a property that can be possessed by things which are desirable, where this in turn is a matter of the differential strengths of the desires that a subject would have if she had a maximally informed and coherent and unified desire set. This means that they only have to explain the remaining two features of evaluative judgements, Certitude and Robustness, in terms of a structural feature of the psychological state that such judgements express. Non-cognitivists, by contrast, must attempt to explain all three features in terms of structural features of the psychological state that such judgements express. Certitude, Importance, and Robustness must one and all reduce, in some yet to be specified way, to structural features of the desires that non-cognitivists tell us we have in so far as we have evaluative commitments.

Here, however, a problem looms. For what structural features do desires possess? As far as I can see, desires possess just two structural features that look like they will be of any use. Desires differ from each other in terms of their strength: a subject's desire that she experiences a certain amount of pleasure might be weaker than, as strong as, or stronger than, her desire that she enjoys a certain degree of autonomy, for example. And the strength of an agent's desires may vary over time under the impact of information and reflection: at one time a subject's desire that she experiences a certain amount of pleasure might be weaker than her desire that she enjoys a certain degree of autonomy, for example, but, under the impact of information and reflection, at a later time her desire that she experiences a certain amount of pleasure might be stronger than her desire that she enjoys a certain degree of autonomy.

Now consider a very flat-footed form of non-cognitivism, just to make the problem that looms vivid, a form of non-cognitivism according to which a subject's judgement that experiencing a certain amount of pleasure is desirable is an expression of her desire that she experiences pleasure, and whose judgement that being autonomous to a certain degree is desirable is an expression of her desire that she enjoys that degree of

autonomy. Since subject's desires can vary in strength, that can represent something, presumably either Importance or Certitude. And since the strength of the subjects' desires over time can vary under the impact of information and reflection then, at least if we allowed strength of desire to represent Certitude, that too can represent something, namely, Robustness. But this leaves at least one thing, either Importance or Certitude, not represented at all. The problem that looms, then, is that a non-cognitivist seems not to have the resources to accommodate all three of the further belief-like features of evaluative judgements that we commonsensically ascribe to them. They can accommodate either Importance, or they can accommodate Certitude and Robustness, but they cannot accommodate all of Importance, Certitude and Robustness.

At this point it might be thought that non-cognitivists should simply insist that we not consider such a flat-footed version of their theory. It is worthwhile considering this response. Instead of holding that an evaluative judgement expresses a first-order desire, what happens if we hold, in the spirit of Blackburn's suggestion that there is a staircase of emotional and practical ascent, that an evaluative judgement expresses (roughly speaking) a second-order desire? Mightn't we then be able to represent both Importance and Certitude, in addition to Robustness, in the following terms?

A subject who judges that it is desirable to experience a certain amount of pleasure, and who also judges that it is desirable to be autonomous to a certain degree, and who judges, as well, that experiencing that amount of pleasure is less desirable than being autonomous to that degree, is someone who (roughly speaking) desires to desire that she experiences that amount of pleasure, and desires to desire that she be autonomous to that extent, and, in addition, the suggestion might be, she is someone whose desired desire that she experiences pleasure is weaker than her desired desire that she be autonomous. Importance, it might be claimed, can in this way be represented by the strength of the desired desire. Certitude, it might then be thought, could be represented by the relative strengths of the second-order desires themselves. In other words, a subject whose desire that she desires that she experiences a certain amount of pleasure is stronger than her desire that she desires that she be autonomous is someone who is more confident about the value of pleasure than she is about the

value of autonomy. Robustness could then be explained as before in terms of the diachronic sensitivity of the second-order desires to information and reflection.

Speaking for myself, I must say that I don't find this less flat-footed version of non-cognitivism very compelling. Why should we suppose that Importance maps onto the strength of the desired desire, and Certitude onto the strength of the second-order desire itself, rather than vice versa? The assignment seems arbitrary, and, for that reason alone, difficult to believe. But there is another, and more striking, problem with the suggestion as well. For it is difficult to see how to square the proposed account of Certitude and Importance with the observation that a subject who is less certain about the value of autonomy and more certain about the value of pleasure may none the less be rationally required to desire to be autonomous more than she desires to experience pleasure because of her assignments of relative desirability.

In order to see why, consider a subject who has a strong desire that she has a weak desire that she experiences pleasure, and a weak desire that she has a strong desire that she be autonomous. According to the proposal under consideration, this is supposed to amount to her having a certain level of confidence that experiencing pleasure is desirable to a certain degree and a lesser level of confidence that being autonomous is desirable to a greater degree. But whereas, intuitively, it should be possible for the relativities in the levels of confidence and desirability to be such that the subject is rationally required to have a stronger first-order desire that she be autonomous and a less strong first-order desire that she experience pleasure, there doesn't seem to be any way for the fact that her desired desire to be autonomous is strong, as compared with her desired desire that she experience pleasure, to have any effect whatsoever on what it is rational for such a subject to first-order desire more. The simple fact is that she wants more strongly to have a weak desire to experience pleasure, which seems to entail that, in cases of conflicts between values, greater confidence will always determine what it is rational for a subject to want most.

The trajectory of the present argument would seem to be plain. As I warned at the outset, the problem highlighted arises because of the paucity of resources available to non-cognitivism. It is therefore irresistible to conclude that not only is non-cognitivism unable to make room for the standard belief-like features of evaluative

judgements — the fact that they enjoy (some analogue of) epistemic support — but that it is also unable to give an adequate account of the way in which an agent's values explain her actions. Commonsense tells us that there are three further belief-like features of evaluative judgements — Certitude, Importance, and Robustness — and that these three features play a crucial role when it comes to understanding the explanatory connection between an agent's values and her actions. When it comes to the explanation of action, less confidence that something is much much more desirable may well trump greater confidence that something is much much less desirable. Yet non-cognitivism seems quite unable to explain why this is so. Desires simply don't have enough features to represent everything that they need to represent. The upshot would thus seem to be that non-cognitivism is unable to accommodate either the belief-like or the desire-like features of evaluative judgements.

Conclusion

As I said at the outset, it is commonly assumed that cognitivists can accommodate the belief-like features of evaluative judgement, and that they confront an insurmountable problem in trying to account for the potential to lead all the way to action that beliefs with that sort of content must have. Non-cognitivists, by contrast, are commonly assumed to be able to accommodate the desire-like features of evaluative judgement, but then to face the insurmountable problem of trying to account for the potential for justification and rational defence that such motivational antecedents must therefore have.

The argument of this paper has been that this common assumption is simply mistaken. There is a form of cognitivism that can easily accommodate both the belief-like and the desire-like features of evaluative judgement, and no form of non-cognitivism, surprisingly, can accommodate neither. The belief-like features of evaluative judgement are complex and these complexities play a crucial role when it comes to understanding both the potential that evaluative judgements have for justification and rational defence and the way in which an agent's values explain her actions. Non-cognitivism is ill-equipped to accommodate these complexities.

To repeat what I said at the outset, my rather sweeping negative claim about non-cognitivism might be wrong, of course. Perhaps there is some unknown form of non-

cognitivism waiting in the wings that can explain either the belief-like features, or the desire-like features, or perhaps even both. But if there is then I suspect that that will be because this more plausible form of non-cognitivism differs in some fundamental way from the versions that are currently on offer. I will be happy enough if the effect of this paper is to force non-cognitivists to come up with a more plausible formulation of their theory.

REFERENCES

Blackburn, Simon 1998: Ruling Passions (Oxford: Clarendon Press).

_____ 2002: 'Replies' in Philosophy and Phenomenological Research. 164-176.

Brink, David O. 1989: Moral Realism and the Foundations of Ethics (New York: Cambridge University Press).

Hare, R.M. 1952: The Language of Morals (Oxford: Oxford University Press).

Humberstone, Lloyd 1987: 'Wanting as Believing' in Canadian Journal of Philosophy, pp.49-62.

Parfit, Derek 1984: Reasons and Persons (Oxford: Oxford University Press).

Pettit, Philip and Michael Smith 2000: 'Global Consequentialism' in Brad Hooker, Elinor Mason and Dale E. Miller, eds, Morality, Rules, and Consequences: A Critical Reader (Edinburgh: Edinburgh University Press). 121-133.

Rawls, John 1951: 'Outline of a Decision Procedure for Ethics', Philosophical Review. 177-97.

Smith, Michael 1994: The Moral Problem (Oxford: Basil Blackwell).

_____ 1995: 'Internal Reasons' in Philosophy and Phenomenological Research, pp.109-131.

_____ 1997. 'In Defense of The Moral Problem: A Reply to Brink, Copp and Sayre-McCord' in Ethics. 84-119.

_____ 2001a: 'The Incoherence Argument: Reply to Schafer-Landau' in Analysis. 254-266.

_____ 2001b: 'Immodest Consequentialism and Character' in Utilitas. 173-194.

_____ 2002a: 'Which Passions Rule?' in Philosophy and Phenomenological Research. 157-63.

_____ 2002b: 'Evaluation, Uncertainty and Motivation' in Ethical Theory and Moral Practice. 305-320.

_____ 2003: 'Rational Capacities' in Sarah Stroud and Christine Tappolet, eds, Weakness of Will and Varieties of Practical Irrationality (Oxford: Oxford University Press). 17-38.

Watson, Gary 1975: 'Free Agency' reprinted in Gary Watson, ed., Free Will (Oxford: Oxford University Press, 1982). 96-110.

Williams, Bernard 1980: 'Internal and External Reasons' reprinted in his Moral Luck (Cambridge: Cambridge University Press, 1981).
