Suppose, just for the sake of argument, that the version of the dispositional theory of value that I myself prefer is correct (see, for example, Smith 1989, 1994a, 1997; compare Lewis 1989, Johnston 1989): when a subject judges it desirable for p to be the case in certain circumstances C, this is a matter of her believing that she would want p to be the case in C if she were in a state that eludes all forms of criticism from the point of view of reason—or, for short, and perhaps somewhat misleadingly (Copp 1997), if she were fully rational.¹

More precisely, if still somewhat misleadingly, let’s suppose that when a subject judges it desirable that p in C this is a matter of her believing that, in those nearby possible worlds in which she is fully rational—let’s call these the ‘evaluating possible worlds’—she wants that, in those possible worlds in which C obtains—let’s call these the ‘evaluated possible worlds’—p obtains.

Once we have supposed this to be so it is, I think, extremely tempting to suppose that we have thereby either explicitly or implicitly taken a stand on certain crucial debates in meta-ethics: tempting to suppose that we must be cognitivists as opposed to non-cognitivists; relativists as opposed to non-relativists; and realists as opposed to irrealists. We must be cognitivists because we have supposed that evaluative judgement is a species of belief. We must be relativists because we have supposed that the truth conditions of a subject’s evaluative beliefs are fixed by whatever that subject would want if she were fully rational, and hence are relative to what that subject herself actually desires: the contents of different subjects’ evaluative beliefs must be different from each other simply because they are beliefs of different subjects. And we must be realists because there is every reason to suppose that there are facts about what different subjects want in possible worlds like the evaluating world, possible worlds in which they are fully rational; every reason to suppose, in other words, that some such beliefs are true. The realism that is thus implied contrasts not with the kind of irrealism defended by non-cognitivists, but with
that defended by cognitivists like John Mackie who thinks that we should all be error theorists about value (Mackie 1997).

Notwithstanding the very understandable temptation to think that all of this is so, I will argue that the implications of the dispositional theory are either different or, at the very least, much less clear. Though the dispositional theory does give us grounds on which to make a case for cognitivism, I will argue that making that case requires that we appeal to certain controversial supplementary premises (§1). As regards relativism, I will argue that the dispositional theory not only has no such implication, but that, on its face, it commits us, if anything, to non-relativism (§2). And as regards realism, I will argue that the dispositional theory leaves it very much an open question whether realism or irrealism is true. That debate, too, turns on the truth of certain supplementary, and highly controversial, premises (§3).

1. The Dispositional Theory and Cognitivism

Since the issue that divides cognitivists from non-cognitivists is, by definition, whether evaluative judgements are expressions of belief (the cognitivists’ view) or some non-belief state, a state of desire or whatever (the non-cognitivists’ view), it may seem inevitable that, having taken it as given that a subject’s judging desirable is a matter of her believing that she would have certain desires if she were fully rational, we must be cognitivists. Inevitable though it might appear, however, it seems to me this line of reasoning is mistaken.

Consider the following, much shorter, argument for cognitivism, by way of comparison.

Premise: When a subject judges it desirable for \( p \) to be the case in certain circumstances \( C \), this is a matter of her believing that it is desirable for \( p \) to be the case in \( C \).

Conclusion: A cognitivist theory of desirability judgements is correct.

Now I take it that no-one will find this argument convincing. The problem is not that the premise is false. The premise, being simply a correct report of the way in which we use the English word ‘belief’, is true. The problem is rather that, precisely because this is why the premise is true, it is too weak to establish the truth of the desired conclusion. Moreover the reason why this is so should be evident. Quite generally, the mere fact that we ordinarily describe things in certain ways does nothing to show that those descriptions apply to those things strictly speaking. The mere fact that we ordinarily describe certain people as ‘pigs’, for example, does nothing to show that they are pigs strictly speaking. It shows, at most, that they are like pigs in certain respects. Likewise, then, the mere fact that we ordinarily describe people as having evaluative beliefs does nothing to show that the attitudes thus described are beliefs
strictly speaking either. It shows, at most, that the attitudes are like beliefs in certain respects. This is something that can be agreed by cognitivists and non-cognitivists alike.

The question that naturally arises is what more we need to establish in order to show that evaluative judgements express beliefs strictly speaking. In order for the attitudes that people have when we ordinarily describe them as having evaluative beliefs to be beliefs strictly speaking, the sentences that we use to give the ‘contents’ of these attitudes—sentences like ‘It is desirable for p to be the case in C’—must be truth-apt. Belief is, after all, the attitude of taking something to be a certain way, and specifying one of the ways that something could be is the distinctive role of a truth-apt sentence (Jackson, Oppy and Smith 1994). It is this that cognitivists and non-cognitivists disagree about. Non-cognitivists think that there is a compelling reason to think that these sentences do not purport to say how things are (Hare 1952; Blackburn 1984). Non-cognitivists insist that those who use these sentences properly must, at least absent practical irrationality, be in some sort of non-cognitive state—a motivational state, or a state of approval—and the only way in which this could be so is if the function of these sentences wasn’t to say how things are, but was rather to express that very non-cognitive state. It follows, at least as the non-cognitivists see things, that though we do not violate any rules of English usage when we describe those who are disposed to make evaluative judgements as having ‘evaluative beliefs’, this must be understood as loose talk (Smith 1994b; Blackburn 1998). The attitudes in question, though like beliefs in certain respects, are not beliefs strictly speaking.

Let’s now return to the original argument. If what we have just said is right then the mere fact that we would ordinarily describe a subject as ‘believing’ that she would want p to be the case in certain circumstances C if she were fully rational does nothing to show that this attitude is a belief strictly speaking either. This too is so only if the sentence we use to give the ‘content’ of this attitude—the sentence ‘Subject S would want p to be the case in circumstances C if she were fully rational’—is truth-apt; in other words, only if the function of the sentence is to specify a way that things could be. Here too cognitivists and non-cognitivists might therefore disagree. Non-cognitivists might insist that the very same consideration that shows that the sentence ‘It is desirable for p to be the case in C’ is not truth-apt shows that the sentence ‘Subject S would want p to be the case in circumstances C if she were fully rational’ is not truth-apt either (Blackburn 1998). In other words, they might argue that the connection between the state that one is in when one is disposed to make that judgement and some sort of non-cognitive state—a motivation, or a state of approval—shows that the role of the sentence is to express that non-cognitive state. If they were right about this we would once again have to conclude that we only speak loosely when we say that subjects believe that they would want p to be the case in C if they were fully rational. Strictly speaking they would not be in a state of belief at all. They would be in a non-cognitive state.
The upshot is thus that, even if we grant that a subject's judging desirable is a matter of her believing that she would have certain desires if she were fully rational, it simply doesn't follow that we thereby commit ourselves to the truth of cognitivism. In order to establish the truth of cognitivism we must establish the truth of further supplementary premises. Specifically, we must establish that the function of the sentence ‘Subject S would want p to be the case in circumstances C if she were fully rational’ is to specify a way that that subject could be. Moreover, in order to do this without begging the question against the non-cognitivists we must establish something else as well, namely, that when subjects have beliefs about themselves being that way, whatever that way is, their beliefs have the kind of connection with non-cognitive attitudes—with motivating attitudes, or attitudes of approval—that non-cognitivists say no belief can have.

Can these supplementary premises be provided? It seems to me that they can, but this is of course all very controversial. As I see things, the claim that a subject is fully rational—where, remember, in the present context this is just to say that the subject is in a state that eludes all forms of criticism from the point of view of reason—entails a set of quite specific claims about the way that that subject is.

For example, following Bernard Williams's lead, we must suppose that the fact that a subject's desires are based on ignorance or error is, at one and the same time, a determinate way that those desires are and, for that very reason, a criticism of those desires from the point of view of reason (Williams 1980). This is because someone who was perfect, from the point of view of reason, would be omniscient and make no mistakes. But, if this is right, then it follows that there is at least one counterexample to the quite general non-cognitivist suggestion that to say that a subject's desires are liable to criticism from the point of view of reason is not to specify a way that those desires are, but is rather to express a desire about, or some other non-belief attitude towards, those desires being a certain way. Indeed, it would seem to be completely irrelevant whether those who use the term 'fully rational' happen to desire people not to have desires that are based on ignorance or error, or whether they happen to have any other non-cognitive attitude towards them. Instead it seems to be analytic that desires based on ignorance and error are liable to criticism from the point of view of reason.

Once we see that this is so an obvious question presents itself. Are there other ways a subject's desires can be which, as such, make those desires criticizable from the point of view of reason? And the answer, as I see things, is that there most certainly are. To say that a subject has a desire set that, as a whole, exhibits incoherence, for example, or to say that she has a desire set which, as a whole, exhibits a lack of unity, is equally a specification of a way that that desire set can be and a criticism of that desire set from the point of view of reason. Again, it would seem to be completely irrelevant whether those who use the term 'fully rational' happen to desire people to have desires that
are coherent and unified, or whether they have any other non-cognitive attitude towards them. Someone who claimed that, according to their usage of the words ‘rational criticism’, to say of a set of desires that they lack coherence and unity isn’t a form of rational criticism is someone who simply doesn’t understand what rational criticism is. It would therefore seem once again to be analytic that a desire set that lacks coherence or unity is, as such, a desire set that is liable to criticism from the point of view of reason.2

How might non-cognitivists try to resist this line of argument? Following a suggestion made by Geoffrey Sayre-McCord, they might profess not to understand what is meant by the terms ‘coherence’ and ‘unity’ when these terms are applied to sets of desires (Sayre-McCord 1997). Alternatively, following a suggestion made to me by Sigrún Svavarsdóttir in conversation, they might agree that they can understand what is meant, but only if the terms ‘coherent and unified’ as applied to sets of desires are taken to mean something like ‘co-satisfiable’, an interpretation which is of little help given that desires that do not form a coherent and unified set in this sense—that is, desires that do not constitute a co-satisfiable set—are hardly, as such, criticizable from the point of view of reason.

But neither of these responses seems to me to be very plausible. When applied to sets of beliefs the terms ‘coherence’ and ‘unity’ are plainly comprehensible as specifying ways that those sets of beliefs can be. The relations that hold between desires when the terms ‘coherence’ and ‘unity’ are applied to them are, as I will go on to argue, plainly relations of exactly the same kind as these. As such it is, I think, hard to take seriously the objections of both those who profess not to understand what is meant by the terms ‘coherence’ and ‘unity’ when these terms are applied to sets of desires, and those who wish to offer an idiosyncratic interpretation of the terms, an interpretation that has nothing to do with the possibility of criticism from the point of view of reason.

In order to see that the terms ‘coherent’ and ‘unified’ really do specify ways that sets of beliefs can be, consider various sets of beliefs that combine, on the one hand, ordinary observational beliefs together with, on the other, beliefs about the behaviour of theoretical entities, theoretical entities whose behaviour is supposed to explain those observations. It is, I take it, completely uncontroversial that some of these sets of beliefs will exhibit more or less in the way of coherence and unity than other sets. Moreover I take it that when we so describe sets of beliefs we plainly specify a way that these beliefs are. We would perhaps have difficulty specifying in terms other than ‘coherent’ and ‘unified’ what that way is—the concepts of coherence and unity are perhaps, in this respect, rather like recognitional concepts—and the classifications might be vague at the borders, but, at least after allowing for these peculiarities, it seems to me that we would have little difficulty in principle in providing an interpersonally agreed ordering of the various sets of beliefs from those that exhibit most in the way of coherence and unity to those that exhibit least. To this extent being coherent and being unified would seem to be ways that these sets of beliefs can be. More-
over, to say that a set of beliefs is a way such that it exhibits less rather than more in the way of coherence and unity is simultaneously a criticism of that set of beliefs from the point of view of reason. It is analytic that, at least other things being equal, a more coherent set of beliefs is less liable to criticism from the point of view of reason, and the same goes for a set of beliefs that is more unified.

If this is right, however, then it seems plain that much the same can be said about sets of evaluative judgements. Consider the variety of sets of judgements that combine what Rawls calls our considered evaluative judgements—these are evaluative judgements about rather specific situations in which we have the greatest confidence—together with various alternative sets of judgements that we might make about general evaluative principles, general evaluative principles which are supposed to justify these considered judgements (Rawls 1951). In this case, too, it seems that we can order the sets from those that exhibit most in the way of coherence and unity to those that exhibit least. There is, in other words, nothing about the nature of the relations that exist between the judgements themselves, whether those judgements are best thought of as expressing beliefs or desires, that requires us to suppose that these relations aren’t of the very same kind as the relations that exist between the sets of beliefs just considered. In this case too, then, it seems that being coherent and being unified specify ways that these sets of judgements can be. Moreover in this case, too, to say of a set of evaluative judgements that it exhibits less rather than more in the way of coherence and unity is simultaneously a criticism of that set of judgements from the point of view of reason. It is analytic that, at least other things being equal, a more coherent set of evaluative judgements is less liable to criticism from the point of view of reason, and the same goes for a set of evaluative judgements that is more unified.

Finally, consider the relations that exist between the various sets of desires that we get if, for each set of evaluative judgements of the kind just mentioned we substitute a specific desire that A Ï€s in circumstances C for each specific evaluative judgement of the form 'It is desirable that A Ï€s in circumstances C', and we substitute a general desire that (x) Ï€s in circumstances C') for each general evaluative judgement expressible in the form ‘(x) (It is desirable that x Ï€s in circumstances C')’. Once we have granted that the various sets of evaluative judgements of the kind just mentioned exhibit relations of coherence and unity, it seems to me that there is no alternative but to suppose that the isomorphic sets of desires just described exhibit those same relations of coherence and unity. In this case, too, it seems that we must suppose that being coherent and unified specify ways that these sets of desires can be. In this case, too, to say of a set of desires that it is a way such that it exhibits less rather than more coherence and unity is simultaneously a criticism of that set of desires from the point of view of reason.

The upshot is that those who deny that claims of the form ‘Subject S would want that p be the case in circumstances C if she were fully rational’ specify a
way that things could be look to be on very shaky ground. To say that S would have certain desires if she were fully rational is to say that she would have those desires if she had a set of desires that eludes all forms of criticism from the point of view of reason, and, so far, we have seen that this entails that her desire set would have to be certain quite specific ways: maximally informed, coherent, and unified. Though we haven’t yet been given any reason to suppose that this provides an exhaustive account of the ways that S’s desire set would have to be to be fully rational, we have so far been given no reason to suppose that such further conditions as we might add wouldn’t simply be further specifications of ways that sets of desires have to be in order for them to count as fully rational. Though this does not constitute a decisive proof of cognitivism, it should at least be agreed that a cognitivist account of subjects’ judgements about what they would want if they were fully rational looks to be on the cards. Pro tem, then, we should suppose that such judgements express not just beliefs loosely speaking, but beliefs strictly speaking.

I said above that if we aren’t to beg the question against the non-cognitivists then, in order to show that evaluative judgements really do express beliefs strictly speaking, we would have to show not just that there is a way that someone takes things to be when they believe that they would desire p to be the case if they were fully rational, but also that their taking things to be that way has the kind of necessary connection with motivation that non-cognitivists insist evaluative judgements have. We must show, in other words, that, absent practical irrationality, a subject who believes that she would desire p to be the case in C if she were fully rational does indeed desire that p in C. Can this argument be given? Though this too is controversial (Shafer-Landau 1999), it seems to me that the argument can indeed be given (Smith 2001).

Imagine a case in which a subject comes to believe that (say) she would desire that she abstains from eating sweets in the circumstances of action that she presently faces if she had a maximally informed and coherent and unified set of desires, but that she doesn’t have any desire at all to abstain. She desires to eat sweets instead. Now consider the pair of psychological states that comprises her belief that she would desire that she abstains from eating sweets in the circumstances of action that she presently faces if she had a maximally informed and coherent and unified set of desires, and which also comprises the desire that she abstains from eating sweets, and compare this pair of psychological states with the pair that comprises her belief that she would desire that she abstains from eating sweets in the circumstances of action that she presently faces if she had a maximally informed and coherent and unified set of desires, but which also comprises instead a desire to eat sweets. Which of these pairs of psychological states is more coherent?

The answer would seem to be plain enough. The first pair is much more coherent than the second. There is disequilibrium or dissonance or failure of fit involved in believing that you would desire yourself to act in a certain way in certain circumstances if you had a maximally informed and coherent and uni-
fied desire set, and yet not desiring to act in that way. The failure to desire to act in that way is, after all, something that you yourself disown; from your perspective it makes no sense, given the rest of your desires; by your own lights it is a state that you would not be in if you were in various ways better than you actually are: more informed, more coherent, more unified in your desiderative outlook. There would therefore seem to be 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 that comprises both the subject’s belief that she would desire that she abstains from eating sweets in the circumstances of action that she presently faces and the desire that she abstains from eating sweets.

If this is right, however, then it follows immediately that if the subject is rational, in the relatively mundane sense of having and exercising a capacity to have the psychological states that coherence demands of her, then, at least abstracting away from such other dynamic changes in her beliefs as might occur for evidential reasons (Arpaly 2000), that subject will end up having a desire that matches her belief about what she would want herself to do if she had a maximally informed and coherent and unified desire set. In other words, in the particular case under discussion, she will end up losing her desire to eat sweets and acquiring a desire to abstain from eating sweets instead. Subjects’ beliefs about what they would want if they were fully rational thus seem both to be beliefs strictly speaking and to be beliefs which have the kind of necessary connection with motivation that non-cognitivists insist evaluative judgements have. Absent practical irrationality—that is to say, absent a failure either to have or to exercise the capacity to have the psychological states that coherence demands of her—a subject who believes that she would want p to be the case in C if she had a maximally informed and coherent and unified desire set will indeed desire p to be the case in C. The non-cognitivists’ reasons for supposing that evaluative judgements are not beliefs strictly speaking therefore seem, in the end, to be unconvincing.

2. The Dispositional Theory and Relativism

Let’s now suppose not just that when a subject judges it desirable that p be the case in certain circumstances, C, this is a matter of her believing that she would want that p be the case in C if she were fully rational, but also that, for the reasons just given (Ol), these states are beliefs strictly speaking. Once we grant this it is, I think, extremely tempting to suppose that we thereby commit ourselves to the truth of relativism (Johnston 1989). Here is why.

Consider two people, A and B, who appear to have the same evaluative belief, the belief that it is desirable that p. Given the equivalence, what A believes is that she, A, would desire that p if she, A, had a set of desires that is maximally informed and coherent and unified. But what B believes, by con-
trast, is that he, B, would desire that p if he, B, had a set of desires that is maximally informed and coherent and unified. A's and B's beliefs thus have quite different truth conditions. A's belief is made true by the desires that she would have, never mind about B's, and B's by the desires he would have, never mind about A's. If this is right, though—that is, if A's and B's beliefs do indeed have different truth conditions—then the appearance that they have the same belief, when they each believe that it is desirable that p, is misleading. A is more accurately represented as believing that it is desirable\(_A\) that p, and B as believing that it is desirable\(_B\) that p. The suggested equivalence thus seems to imply the truth of relativism.

Tempting though this line of thought is, we should, I think, resist it. In order to see why, it will be helpful to work through an example of a particular substituend for 'p' in the belief that it is desirable that p. To anticipate, my argument will be that once we pay due regard to three facts about values—the fact that value is universalizable, the fact that value can be either neutral or egocentric, and the fact that some neutral values and egocentric values are commensurable—it becomes plain that evaluations not only are not, but that they could not be, relative in the way that has just been suggested. If anything, these considerations suggest that we are committed to a non-relativist conception of value.

Imagine someone, A, with a belief whose content we might initially think she should express in the following sentence:

\[\text{It is desirable that my}_A\text{ children fare well.}\]

where the subscript to the 'my' simply serves to make it explicit whose children are being referred to. Given the equivalence, this amounts to A's having a belief whose content she could express in the following sentence:

\[\text{I}_A\text{ would want that my}_A\text{ children fare well if I}_A\text{ had a set of desires that is maximally informed and coherent and unified.}\]

Complications immediately arise, however.

It is, after all, a conceptual truth that evaluations are one and all universalizable. It therefore follows that specific evaluations, like this one, must be derivable from more universal beliefs to which agents are committed. We must therefore ask which universal evaluation A is committed to simply in virtue of having the belief that it is desirable that her own children fare well. In other words—and, remember, we are supposing equivalently—we must ask which universal evaluation A is committed to simply in virtue of having the belief that she would want that her children fare well if she had a set of desires that is maximally informed and coherent and unified.

One possibility is that A is committed to a universal evaluative belief with the following content:
(x) (It is desirable that x’s children fare well)

or, perhaps equivalently:

(x) (IA would want that x’s children fare well if IA had a set of desires that is maximally informed and coherent and unified)

But while this might be one possibility, it certainly isn’t the only possibility. It simply assumes that A assigns neutral, or non-egocentric, value to the welfare of people’s children—assumes, in other words, that A believes it equally desirable that her own children fare well and that other people’s children fare well; assumes, in terms of the equivalence, that she would desire equally that her own children fare well and that other people’s children fare well if she had a set of desires that is maximally informed and coherent and unified—whereas the original belief is plainly ambiguous between that possibility and the quite different possibility that she assigns egocentric value to the welfare of her children, and hence that the desires she would have if she had a set of desires that is maximally informed and coherent and unified would be quite specifically desires about the welfare of her own children. This, in turn, suggests that we went wrong in trying to give the content of her original belief. We should have noted that that belief is ambiguous, and insisted that it be disambiguated before we give its equivalent. Let’s therefore start again.

The content of A’s original belief, the belief that it is desirable that her children fare well, is ambiguous. What she has is either a belief the content of which she could express in the following sentence:

It is desirable that my children fare well

—this is what she believes if she assigns the welfare of her children egocentric value—or, alternatively, it might be suggested, she has a belief the content of which she could best express in the following sentence:

It is desirable that my children fare well

which is what she believes if she assigns the welfare of her children neutral value. Because evaluations are one and all universalizable, these beliefs might then be thought to commit A to universal beliefs that she could best express in one or another of the following sentences:

(x) (It is desirable that x’s children fare well)

—this is the content of the universal belief to which she is committed if she assigns the welfare of her children egocentric value—or, alternatively,
Exploring the Implications of the Dispositional Theory of Value

(x) (It is desirable that x’s children fare well)

which is the content of the universal belief to which she is committed if she assigns the welfare of her children neutral value.

But this can’t be quite right either. If it were it would follow, implausibly, that such assignments of neutral and egocentric value to the welfare of children are radically incommensurable. Neutral value would, after all, be a completely different property from egocentric value—egocentric value would be an indexed property, whereas neutral value an unindexed property—and this in turn would mean that we couldn’t sensibly ask someone who assigns both neutral value to the welfare of people’s children and egocentric value to the welfare of her own children whether the egocentric value that she assigns to her own children’s welfare was greater than or less than the neutral value that she assigns to their welfare. The comparative concept of desirability would, after all, have to be the comparative form of either the indexed property or the unindexed property. We could only ask A questions such as whether her own children’s welfare has more egocentric value than the egocentric value possessed by other people’s children (a question the answer to which is that it plainly does, since other people’s children have no such egocentric value), or whether the welfare of other people’s children has more neutral value than that possessed by her own children (the answer to which is plainly that it doesn’t, since other people’s children’s welfare has the same neutral value as that possessed by her own).

As I said, it seems to me that the idea that assignments of neutral and egocentric value are radically incommensurable in this way is manifestly implausible. Those who assign neutral value to the welfare of people’s children and egocentric value to the welfare of their own children have no problem at all comparing these two values. Indeed, I think that most people would insist that the egocentric value that they assign to their own children’s welfare is greater than the neutral value they assign to their own children’s welfare, and to the welfare of the children of others. This is why they feel totally justified in giving benefits to their own children over comparable benefits to strangers. We are, however, yet to find an account of the logical form of evaluations that makes it plain just what it might mean when we make such comparative evaluative claims.

Returning to the example we have been working through so far, what this suggests, I think, is that A should express the content of her original belief in the following sentence:

It is desirable_A that my_A children fare well

and that in order to disambiguate this belief we must say whether it, in turn, is derived from a universal belief with the following content:

(x) (It is desirable_x that x’s children fare well)
—this is the content of the universal belief to which A is committed if she assigns the welfare of her children egocentric value—or, alternatively, in a universal belief with the following content:

\[ (x)(y) \text{ (It is desirable}_x \text{ that } y's \text{ children fare well) } \]

This is the content of the universal belief to which she is committed if she assigns the welfare of her children neutral value. In other words, and rather naturally I think, the difference between a neutral evaluation of the welfare of a subject’s children and an egocentric evaluation is that whereas the egocentric evaluation is an evaluation from that subject’s own point of view, the neutral evaluation is an evaluation from everyone’s point of view. Moreover, though the subscript on the desirability predicate in the neutral evaluation might look completely idle when neutral evaluations are considered in isolation from egocentric evaluations, the fact that there is such an index on the desirability predicate in the neutral evaluation is absolutely crucial when it comes to an understanding of how comparisons of neutral and egocentric value are possible. Much as I suggested above, for example, A might be committed to a comparative universal evaluative belief with the following content:

\[ (x)(y) \text{ ((It is desirable}_x \text{ that } x's \text{ children fare well}) \& (It is desirable}_x \text{ that } y's \text{ children fare well}) \& (It is more desirable}_x \text{ that } x's \text{ children fare well than that } y's \text{ children fare well}) \]

This might be why the egocentric value that A assigns to the welfare of her own children is greater than the neutral value that she assigns to the welfare of people's children quite generally, including even her own children.

Note what we have done so far. So far we have simply focussed on evaluations themselves—particular substituends for the ‘p’ in the proposition ‘It is desirable that p’—and asked how, in the light of three facts about values—the facts that evaluations are universalizable, that evaluations can be assignments of either neutral value or assignments of egocentric value, and that it is at least possible for neutral value and egocentric value to commensurate—a particular subject, A, should express the content of the evaluations to which she is committed when she believes that it is desirable that her own children fare well. We are now in a position to ask what the contents of A’s beliefs are, given the equivalence postulated by the dispositional theory of value.

We have seen that A’s belief that it is desirable that her own children fare well is ambiguous. In terms of the equivalence, how should she express the content of the universal evaluative beliefs to which she is committed under the various disambiguations? Disambiguating in favour of the possibility that she assigns her children’s welfare egocentric value, it turns out that she is committed to a universal belief with the following content:

\[ (x) \text{ (It is desirable}_x \text{ that } x's \text{ children fare well) } \]
which, given the equivalence, suggests that she is committed to a belief with the following content:

\[(x) \ (x \text{ would want that } x\text{'s children fare well if } x \text{ had a set of desires that is maximally informed and coherent and unified})\]

Alternatively, disambiguating in favour of the possibility that she assigns her children's welfare neutral value, it turns out that she is committed to a universal belief with the following content:

\[(x)(y) \ (\text{It is desirable}_x \text{ that } y\text{'s children fare well})\]

which, given the equivalence, means that she is committed to a belief with the following content:

\[(x)(y) \ (x \text{ would want that } y\text{'s children fare well if } x \text{ had a set of desires that is maximally informed and coherent and unified})\]

A subject committed to a comparative evaluative belief with the following content:

\[(x)(y) \ ((\text{It is desirable}_x \text{ that } x\text{'s children fare well}) \& (\text{It is desirable}_x \text{ that } y\text{'s children fare well}) \& (\text{It is more desirable}_x \text{ that } x\text{'s children fare well than that } y\text{'s children fare well}))\]

is, given the equivalence, committed to a comparative belief with the following content:

\[(x)(y) \ ((x \text{ would want that } x\text{'s children fare well if } x \text{ had a set of desires that is maximally informed and coherent and unified}) \& (x \text{ would want that } y\text{'s children fare well if } x \text{ had a set of desires that is maximally informed and coherent and unified}) \& (\text{the desire } x \text{ has that } x\text{'s children fare well would be stronger than the desire } x \text{ has that } y\text{'s children fare well}))\]

As is perhaps already plain, when it comes to the issue of relativism, the conclusion is therefore exactly the opposite of the one that we were tempted by at the outset.

Imagine, once again, two people, A and B, both of whom believe that it is desirable that their own children fare well. Do their beliefs have the same truth conditions or different truth conditions? Since, as we have seen, A's and B's beliefs are ambiguous, the truth conditions of their beliefs depend on how we disambiguate them. Contrary to the suggestion made at the outset, however, and notwithstanding the fact that A's and B's beliefs are ambiguous, it turns out that, so long as we disambiguate them in the same way, their beliefs have the
very same truth conditions. For, given universalizability, they must either both be committed to a belief with the following content:

\[(x) \text{ (} x \text{ would want that } x \text{'s children fare well if } x \text{ had a set of desires that is maximally informed and coherent and unified)\]

—this is the belief to which they are committed if their original beliefs were about the egocentric value of their children’s welfare—or, alternatively, to a belief with the following content:

\[(x)(y) \text{ (} x \text{ would want that } y \text{'s children fare well if } x \text{ had a set of desires that is maximally informed and coherent and unified)\]

—this is the belief to which they are committed if their original beliefs were about the neutral value of their children’s welfare.

On the assumption that we can generalize on the basis of this example, it would thus seem to follow that, contrary to the suggestion made at the outset, one subject’s evaluative beliefs are made true not just by the desires that she would have if she had a set of desires that was maximally informed and coherent and unified, but also by the desires that every other subject would have if they had a set of desires that was maximally informed and coherent and unified. Far from the suggested equivalence committing us to relativism, then, it appears that it commits us if anything to non-relativism. The truth of a subject’s evaluative beliefs requires that all subjects converge in the desires they would have if they had a set of desires that was maximally informed and coherent and unified.~

3. The Dispositional Theory and Realism

I said at the outset that the dispositional theory would seem to commit us not just to relativism, but also to realism. The connection between these two commitments should be plain.

We suggested initially that a subject’s judgements about the desirability of p’s being the case in circumstances C were made true by whether or not that subject herself would want p to be the case in C if she had a set of desires that was maximally informed and coherent and unified, never mind about what other subjects would want if they had a set of desires that was maximally informed and coherent and unified. If this were right then, since it is so plausible to think that there are some things that a particular subject would want if she had such a set of desires—this only requires that we be able to give determinate content to the relevant counterfactuals, after all—it follows that it would likewise be plausible to suppose that, since some such judgements are true, realism must be true. In other words, we could reject the possibility of an error theory of the kind argued for by John Mackie (1977). However, now that we have seen that
the dispositional theory commits us not to relativism, but, if anything, to the rejection of relativism, we must reevaluate this commitment to realism.

If a subject’s judgements about the desirability of p’s being the case in certain circumstances C are made true not just by whether or not that subject herself would want that p be the case in C if she had a set of desires that was maximally informed and coherent and unified, but also by whether everyone else would want that p in C if they had a set of desires that was maximally informed and coherent and unified, then should we suppose that some such judgements are true? In other words, is it plausible to suppose that there are some desires that all subjects would converge upon if they had desire sets that are maximally informed and coherent and unified? Many will insist that that there are no such desires (Sobel 1999). If they are right then we must conclude that realism is false. The dispositional theory, since it entails non-relativism, entails irrealism. My own view, however, is that this is all far too quick. It is unclear whether there are any desires that all subjects would converge upon if they had desire sets that are maximally informed and coherent and unified, but it is equally unclear that there are no such desires. It therefore seems to me best to suppose that the debate between realists and irrealists is yet to be resolved.

Those who think that the dispositional theory entails irrealism are, I think, impressed by a chain of reasoning much like the following. We can surely imagine two subjects, D and E, each of whom, in actuality, has a single intrinsic desire—that is, a desire that they haven’t derived from some further desire that they have plus a belief about means—but a different one: let’s suppose, for example, that D has an intrinsic desire that p, whereas E an intrinsic desire that q. Furthermore, since their respective desires are intrinsic, we can also imagine that each of them would retain their single intrinsic desire no matter what further information they acquired. But since there is no reason to suppose that the acquisition of any further information would lead D and E to acquire additional intrinsic desires, and since D’s and E’s desire sets, comprising as they do just one desire each, are already as coherent and unified as they could possibly be, it follows that D and E would not converge in their intrinsic desires even if they did have a maximally informed and coherent and unified desire set. They would still diverge. Indeed, they would still just have their respective intrinsic desires that p and that q. It therefore follows that there are no desires that everyone would converge upon if they had a maximally informed and coherent and unified desire set. The hypothetical D and E constitute the decisive counterexample. If realism requires such convergence, then realism is false.

What is wrong with this chain of reasoning? The problem lies in the premise that there is no reason to suppose that the acquisition of further information would lead D and E to acquire any additional intrinsic desires beyond their respective intrinsic desires that p and that q. To begin, let’s be clear what this premise says. It says, inter alia, that there is no information that D and E could acquire such that, having acquired that information, they would be rationally required to acquire intrinsic desires beyond their respective intrinsic desires that
p and that q. But it seems to me that the only reason we would have to accept this premise is if we were to assume, quite generally, that there are no rational principles of the following form:

It is rationally required that subjects who believe that r either give up their belief that r or acquire an intrinsic desire that s.

Yet, as we have already seen, this quite general assumption is false.

This is, in effect, what we discovered earlier when we saw that coherence requires subjects who believe that they would want p to be the case in circumstances C if they had a desire set that was maximally informed and coherent and unified to desire that p be the case in C (§1). We saw, in other words, that the following—an instance of the rational principle we would have to assume quite generally to be false—is in fact true:

It is rationally required that an agent who believes that she would have an intrinsic desire that p be the case in circumstances C if she had a maximally informed and coherent and unified desire set either gives up her belief or acquires an intrinsic desire that p be the case in C.

Now, to be sure, this particular claim connecting the acquisition of information with intrinsic desiring is not sufficient all by itself to show that D and E would converge in their desires. But nor is that required at this stage of the argument. All that is required is that we show what is wrong with the chain of reasoning described above that purports to prove that two hypothetical subjects, D and E, would not converge in their intrinsic desires. The crucial point, to repeat, is that once we see that the quite general assumption that there are no rational principles of the form 'It is rationally required that an agent who believes that r either gives up his belief that r or acquires an intrinsic desire that s' is false—once we remind ourselves, in other words, that reflection can lead us to accept the surprising conclusion that certain instances of that principle are in fact true—then it is hard to see what, beyond dogmatic commitment, would lead anyone to think that further reflection won't lead us to the surprising conclusion that more instances of that principle are true (compare Korsgaard 1986). In particular, it is hard to see what, beyond dogmatic commitment, would lead anyone to think that further reflection won't lead us to discover that further instances of that principle, instances sufficient to show that D and E would converge in their desires, are true.6

At this stage it therefore seems to me that we would be wise to suspend judgement on the debate over realism versus anti-realism. Perhaps further reflection will reveal that such further instances of rational principles of the form 'It is rationally required that an agent who believes that r either gives up his belief that r or acquires an intrinsic desire that s' as are required in order to undergird a convergence in the desires of subjects with a maximally informed
and coherent and unified desire set are true; perhaps it will not. We have little choice but to do the required reflection and see.7

Notes

1. This is somewhat misleading because whereas non-culpable ignorance plausibly constitutes a failure to achieve an ideal of reason, and so something that is in this sense criticizable from that point of view, it may not constitute what we would ordinarily call a failure of rationality. For reasons that will become plain shortly I will, however, ignore these differences of meaning in what follows.

2. It should now be plain why we were right to ignore the fact that the shorthand term ‘being fully rational’ does not mean exactly the same as ‘being in a state that eludes all forms of criticism from the point of view of reason’ (see footnote 1 above). All that is crucial is that we can give an account of the way that subjects are when they are in the latter state. The term ‘being fully rational’ really is just convenient, if misleading, shorthand.

3. I have just said that a subject who believes that she would desire that she acts in a certain way if she had a maximally informed and coherent and unified desire set, but who does not desire to act in that way, is in a state that she would not be in if she were in various ways better than she actually is: more informed, more coherent, more unified in her desiderative outlook. It is important to note that this use of the term ‘better’ trades on an understanding of value that cannot be analysed in the way suggested by the dispositional theory, an understanding according to which it is simply analytic that a subject with a maximally informed and coherent and unified desire set is as good as she can be, and that subjects with desire sets that fall ever shorter of being maximally informed and coherent and unified are subjects who are correspondingly less good. This should perhaps come as no surprise given that being good as can be, in this sense, is simply a matter of being perfect from the point of view of reason. Hallvard Lillehammer suggests that dispositional theories that define value in terms of such a non-dispositional conception of the good give up on any claim to metaphysical modesty (Lillehammer 2000). For more on this see footnote 6 below.

4. Another way of putting the conclusion just reached is that it is in the nature of desires that they are psychological states that are rationally sensitive to our beliefs about what we would desire that we do if we had a set of desires that was maximally informed and coherent and unified, ‘rationally sensitive’ in the sense of being psychological states that we would acquire in the light of such beliefs given that we have a capacity to have the psychological states that coherence demands of us. As we will see later (§3), this is an important conclusion to draw, not just because it undermines one of the main arguments for non-cognitivism, but also because it undermines one of the main arguments for relativism.

5. Though I have argued for this conclusion elsewhere (Smith 1994a, 1997, 1999), note that the argument given in the text is completely new. To repeat, the argument given in the text is that the natural interpretation of the dispositional theory, given three facts about values—the fact that values are universalizable, that values can be either neutral or egocentric, and that at least some neutral values and some egocentric values are commensurable—is a non-relativist interpretation. The argument I have given in the past for the conclusion that we must give a non-relative interpretation of the
dispositional theory has been that only so can we capture the non-arbitrariness of values. The argument given in the text might well prompt the question whether it is so much as possible to formulate a relativist version of the dispositional theory. Would any such formulation have to assume, implausibly, either that values are not universalizable, or that values cannot be both neutral and egocentric, or that neutral values and egocentric values are radically incommensurable? Though I will not spell out the formulation here, let me say, for the record, that I do not think that this is so. It seems to me that it is possible to formulate a relativist version of the dispositional theory, albeit a very unintuitive and ad hoc version, that is consistent with the three claims about values just mentioned. Unsurprisingly, however, it also seems to me that when we spell out the relativist version of the dispositional theory that is consistent with these three claims it becomes manifest just how arbitrary value is, on such a relativist conception.

6. Hallvard Lillehammer (2000) suggests that dispositional theories that define value in terms of a non-dispositional conception of the good give up on any claim to metaphysical modesty. But there would seem to be nothing metaphysically immodest about the claims about the good made in footnote 3 above, and nor would there seem to be anything metaphysically immodest about principles of reason of the form ‘It is rationally required that an agent who believes that r either gives up his belief that r or acquires an intrinsic desire that s’. What is true, of course, is that we do indeed find it surprising that there are true instances of a general principle of that form. But, as the argument given in §2 illustrates, perhaps the real surprise lies in the fact that the argument for that conclusion relies on such uncontroversial premises. It remains to be seen whether the arguments given for any further instances that we might discover rely on such similarly uncontroversial premises.

7. I would like to thank John Broome, David Estlund and Philip Pettit for helpful conversations while I was writing this paper.

References