

PARFIT'S MISTAKEN METAETHICS

Michael Smith

In the first volume of *On What Matters*, Derek Parfit defends a distinctive metaethical view, a view that specifies the relationships he sees between reasons, facts, beliefs, desires, aims, actions, and values. Parfit's metaethical view is a version of what he calls 'Objectivism', and he explicitly argues not just for his preferred version of Objectivism, but against all versions of Subjectivism, the metaethical view with which Objectivism is contrasted. Parfit's arguments for Objectivism, and his criticisms of Subjectivism, seem to me to be mistaken, so my aim is to identify where the mistakes lie. Though I do not expect Parfit to agree with me that the mistakes I identify are mistakes, I proceed on the assumption that this does not matter. The best we can do when we engage with philosophers like Parfit, whose detailed metaethical views are part and parcel of a comprehensive worldview with mutually supporting parts, is to state very carefully where our fundamental disagreements with them lie.

1. *Reasons Fundamentalism and reasons for belief*

At the very beginning of *On What Matters*, Parfit tells us:

We can have reasons to believe something, to do something, to have some desire or aim, and to have many other attitudes and emotions, such as fear, regret, and hope. Reasons are given by facts, such as the fact that someone's finger-prints are on some gun, or that calling an ambulance would save someone's life.

It is hard to explain the *concept* of a reason, or what the phrase 'a reason' means. Facts give us reasons, we might say, when they count in favour of our having some attitude, or our acting in some way. But 'counts in favour of' means roughly 'gives a reason for'. Like some other fundamental concepts, such as those involved in our thoughts about time, consciousness, and possibility, the concept of a reason is indefinable in the sense that it cannot be helpfully explained merely by using words. We must explain such concepts in a different way, by getting people to think thoughts that use these concepts. One example is the thought that we always have a reason to want to avoid being in agony. (Parfit 2011, p.31)

This is Parfit's first introduction of the underlying idea he defends and relies on throughout his discussion of metaethics in *On What Matters*, an idea that we might call 'Reasons Fundamentalism'.

According to Reasons Fundamentalism, there is a fundamental metaphysical relation that holds between facts, on the one hand, and beliefs, desires, aims, and actions, on the other. This relation holds when the fact in question gives a reason for the belief, desire, aim, or action. The relation is metaphysically fundamental, or so the suggestion goes, because we cannot define what it is for a fact to give a reason for some belief, desire, aim, or action, in terms of other facts; we cannot analyze the concept of a fact's giving a reason for some belief, desire, aim, or action, in terms of other concepts; we cannot reduce the property of giving a reason for some belief, desire, aim, or action, to some other property; and so on (in what follows I will not fuss about the differences

between real definition, conceptual analysis, and metaphysical reduction). In particular, we cannot analyze the reasons relation in terms of subjective psychological concepts. Reasons Fundamentalism is thus Parfit's preferred version of Objectivism.

Reasons Fundamentalism provides the foundation for all of Parfit's metaethical views. But how plausible is the idea that the relation of being a reason for is unanalyzable? Focus on reasons for belief. Parfit tells us that the reasons relation that holds between a fact that provides a reason for believing and belief is unanalyzable. But I doubt very much that that is so. Though the idiom isn't Hume's, we can understand Hume's famous endorsement of *deductive* reasoning, and attack on *inductive* reasoning, as tantamount to the suggestion that we can analyze the reasons relation in the case of belief, and hence the concept of reasoning as well, insofar as reasoning is a matter of updating our beliefs. The analysis, according to Hume, is to be given in terms of, *inter alia*, the concept of *entailment* (Hume 1740). The fact that *p* provides a reason for believing a proposition *q*, Hume seems to have thought, if and only if and because *p* entails *q*. Correspondingly, someone engages in reasoning to the extent that he treats propositions as though they stand in entailment relations to each other.

Hume's view is, of course, much maligned, but what's maligned isn't the background assumption that we can analyze the concept of a fact's being a reason for belief in other terms. What's maligned is rather his restrictive tying of the concept of a reason for belief to the concept of entailment; his elevation of deductive reasoning over inductive reasoning. If we reject this Humean idea, and extend our concept of a reason to include cases where certain facts provide evidence for the truth of other propositions, without entailing them, then we might suppose instead that the analysis of a reason is to be given in terms of some more liberal theory about the nature of belief that specifies the conditions under which even inductively formed beliefs count as knowledge. Reasons for beliefs would then be the considerations that provide the justifications for our beliefs in such cases.

What would such a theory look like? It would be a theory in which our concept of a reason for belief is our concept of evidence for the truth of what we believe when our beliefs count as knowledge. Such a theory would presumably draw on the theory of perception, on the one hand, and the theory of epistemic rationality, on the other, and drawing on the latter might in turn require us to draw on confirmation theory, the theory of statistical reasoning, the probability calculus, and so on. Here is not the place to go into details, and, to be perfectly honest, I am not sure how the details should be given. But the very fact that we have Hume's own reductive theory of reasons for belief in terms of entailment as a model, together with the fact that we can specify the materials in terms of which an alternative to Hume's theory might be given, makes the idea that we should embrace Reasons Fundamentalism, at least in the case of reasons for belief, premature to say the least. We should proceed on the assumption that the concept of a reason for belief can be analyzed.

This conclusion has far-reaching consequences. Parfit gives us no reason at all to think that the concept of a reason, in the case of reasons for belief, is unanalyzable, and since we have good reason to proceed on the assumption that the concept can be analyzed, this provides us with excellent reasons to reject Reasons Fundamentalism more generally. For how could the concept of a reason for a desire, or an aim, or an action, be

metaphysically fundamental, if the concept of a reason for belief is not? That could be so only if the concept of a reason were itself a disjunction of something fundamental and something non-fundamental, where the disjuncts have nothing whatsoever in common with each other. The concept of a reason would be a ragbag, like the concept of a number-or-a-dog. On the plausible assumption that our concept of a reason isn't such a ragbag, the only conclusion to draw, at least *pro tem*, is that Reasons Fundamentalism is quite generally false. The concept of reason is the concept of a reason for belief, a concept that we can analyze. The only questions left to ask are how many reason concepts we can define in terms of the concept of a reason for belief, and what exactly those analyses are.

This is how Hume saw things. Armed with a reductive conception of reasons for belief in terms of entailment, he thought that we can give a straightforward analysis of the concept of a reason for having certain of our desires. The concept of a reason for an instrumental desire, Hume thought, reduces to the concept of a reason for belief because instrumental desires are complex psychological states that have means-end beliefs and non-instrumental desires as parts. The concept of a reason for an instrumental desire is therefore just the concept of a reason for the means-end belief part. But since non-instrumental desires do not have beliefs as parts—Hume thought that belief and non-instrumental desires were distinct existences (more on this in §4)—it follows that there is no concept of a reason for having non-instrumental desires. Of course, we might have qualms about Hume's line of reasoning. In particular, we might wonder whether we should follow him in supposing that having a reason to have the belief part of an instrumental desire suffices for having a reason for the instrumental desire itself, given that it has another part, a non-instrumental desire, for which there is no reason at all. But his basic philosophical instincts seem to be right. Either we can reduce the concept of a reason for desire to a reason for belief or the concept of a reason has no application to desires.

For his part, Parfit thinks that Hume's view is plainly wrong. It is wrong, Parfit thinks, because it is inconsistent with the obvious truth that "we always have a reason to want to avoid being in agony". But while it may well be true that we always have a reason to want to avoid being in agony, if the concept of a reason in play isn't a ragbag, then that concept must somehow reduce to the concept of a reason for belief. The question to ask is how that reduction might go.

2. Reasons for action

The question on the table is whether we can reduce the concept of a reason for desire—both instrumental desire and non-instrumental desire—and the concept of a reason for aims, and actions as well, to the concept of a reason for belief. This question will prove much more tractable if we start by asking a more modest question. Can we reduce any of these reason concepts to other concepts?

Though in the passage quoted earlier, Parfit seems to suggest that the concept of a reason for action is just like the concept of a reason for belief, desire, or attitude, and so fundamental, a little further in *On What Matters*, we seem to be told that this is not so.

According to Objectivists, though many reasons for acting can be claimed to be given by the fact that some act would achieve one of our aims, these reasons

derive their force from the facts that give us reasons to *have* these aims. (Parfit 2011, p.45)

Thus, for example, though my reason for typing these words is the fact that typing them will achieve my aim of writing a paper about *On What Matters*, Parfit thinks that this reason derives its force from my reason to have the aim of writing a paper about *On What Matters*. If there were no reason for me to have this aim, I would have no reason to type these words. But since every action is performed with some aim, and since in every such case the force of the reason provided by the fact that the action would achieve the aim derives from the reason to have the aim, it follows that, at least on the assumption that no other reasons for action are in the offing—more on this presently—reasons for action are not fundamental. The concept of a reason for action is just the concept of a reason for having the aim of performing that action. Parfit has "helpfully explained merely by using words" what the concept of a reason for action is after all.

Nor should we be surprised by this result. Parfit's Reasons Fundamentalism shares much in common with Scanlon's Reasons Fundamentalism, and Scanlon says quite explicitly in *What We Owe To Each Other* that the class of things for which "for which reasons can sensibly be asked or offered" is a sub-class of mental states that he calls the "judgement-sensitive attitudes", and that there are therefore reasons for action only because reasons for action are analyzable in terms of reasons for judgement-sensitive attitudes (Scanlon 1998, p.20). Here are the relevant passages.

Scanlon tells us that judgement-sensitive attitudes are those

...that an ideally rational person would come to have whenever that person judged there to be sufficient reasons for them, and that would, in an ideally rational person, 'extinguish' when that person judged them not to be supported by reasons of the appropriate kind. (Scanlon 1998, p.20)

These include intention, hope, fear, admiration, respect, contempt, indignation, and desire (Scanlon 1998, pp.20-21). Since actions are not a kind of attitude, he wonders whether this casts doubt on the idea that the class of judgement-sensitive attitudes is the class of things for which reasons can sensibly be asked and offered, and he answers as follows:

A reason for doing something is almost always a reason for doing it intentionally, so 'reason for action' is not to be contrasted with 'reason for intending'. The connection to action, which is essential to intentions, determines the kinds of reasons that are appropriate for them, but it is the connection with judgment-sensitive attitudes that makes events actions, and hence the kind of things for which reasons can sensibly be asked for and offered at all. (Scanlon 1998, p.21)

In other words, though there are reasons for actions, that is only because the concept of a reason for action reduces to the concept of a reason for having an intention. Assuming that the states Scanlon calls 'intentions' are just what Parfit calls 'aims', it therefore follows that Scanlon and Parfit think alike. They both think that reasons for actions reduce to, or are analyzable in terms of, reasons for aims. Reasons for action are not fundamental for either Parfit or Scanlon.

It is worth pausing to say something about the exceptions Scanlon hints at in saying that a reason for doing something is "almost always" a reason for doing it

intentionally. What are the cases in which a reason for doing something isn't a reason for doing it intentionally, and hence isn't a reason for having some aim? I take it that the cases he has in mind are those in which we have reasons for (say) spontaneously squawking like a chicken. The fact that it would be amusing to spontaneously squawk like a chicken in certain circumstances is, let's suppose, a reason for spontaneously squawking like a chicken in those circumstances. But what's important about this reason is that, contrary to what's just been said, it cannot be reduced to a reason to intend to spontaneously squawk like a chicken, for the simple reason that it is impossible to intend to do something spontaneously. Intentions are states whose role is to lead us to do the thing that figures in their content. Since no one could spontaneously squawk like a chicken by acting on their intention to spontaneously squawk like a chicken, as the content of the intention would undermine the spontaneity of the act, it follows that when one spontaneously squawks like a chicken, though one must have some intention or other, the spontaneity has to appear in the acquisition or execution of the intention, not in its content.

These exceptions show something important about our concept of a reason for action. They show that we must distinguish very sharply between two cases in which one has reasons for action. In one case, reasons for action reduce to reasons to desire to act in certain ways, where you could come to act in those ways because you have the relevant desire. Most cases are like this, and cases in which reasons for action reduce to reasons for having intentions are, we might now suppose, a sub-class of these: intentions are just a sub-class of such desires. My reason for typing these words is a case in point, as my reason for typing these words reduces to my reason for desiring to write a paper about *On What Matters*, and what I desire to do, in this case, is something that I could do because I have this very desire. I could type these words, so writing a paper about *On What Matters*, because I desire to write a paper about *On What Matters*. The desire either is, or is a component of, the intention.

In the alternative kind of case, however, reasons for action reduce to reasons to desire to act in certain ways, but it isn't possible for you to come to act in those ways by acting on the relevant desire. The reason you have to spontaneously squawk like a chicken of this kind. The fact it would be amusing is, let's agree, a reason for having a *desire* that one spontaneously squawks like a chicken—one can, after all, desire that one spontaneously squawks like a chicken, notwithstanding the fact that one cannot intend to do so—but, given that spontaneity figures in the content of this desire, it isn't the case that one could come to spontaneously squawk like a chicken by acting on this very desire. So though most reasons for action reduce to reasons for intending, which are just desires of a certain kind, some reasons for action reduce reasons for desiring where the desires are not of that kind. We will return to this idea again shortly (§3), where it will emerge that reasons for action quite generally reduce to facts about the *desirability* of acting in the relevant ways.

Let me sum up. Though Parfit is right that we talk about "reasons to believe something, to do something, to have some desire or aim, and to have many other attitudes and emotions", even he admits that only some of this talk is talk about the fundamental relation of being a reason for. Some of it is not. Our talk of reasons for action is a case in point, as Parfit thinks that the concept of a reason for action can be analyzed in terms

of the concept of reasons for desires and aims. But once we have this conclusion firmly before our minds, a good question to ask is whether we can do for reasons for desires, aims, and other attitudes, what we have just seen can be done for reasons for action. Just as reasons for action reduce to reasons for desires and aims, might reasons for desires and aims reduce to reasons for belief? If so, then on the assumption that we can fully reduce the concept of a reason for belief in the way suggested at the outset, the upshot would be that we can fully reduce the concept of a reason.

3. *Reasons for desires, aims, and other attitudes*

The question on the table this time is whether we can reduce the concept of a reason for desire, aims, and other attitudes, to the concept of a reason for belief. Judith Jarvis Thomson offers what seems to me to be most promising answer to this question in her *Normativity* (2008).

Thomson's answer starts with an observation about the correctness conditions of a variety of mental states. Many different mental states have *correctness* conditions, Thomson notes, where these conditions can be informally understood as those in which the mental states in question are *deserved* (Thomson 2008, p.116). Not all kinds of mental states can be deserved. There is, for example, no circumstance in which feeling dizzy can be deserved, so feeling dizzy has no correctness condition. But mental states of other kinds can be deserved, so they do have correctness conditions. Thomson illustrates this idea by focusing on the case of belief.

A believing is a correct believing just in case its propositional content is true.

Thus a believing is not marked as a correct believing by the fact that the believer's other beliefs lend weight to, or even entail, the propositional content of his believing. Smith's believing P is not marked as a correct believing by the fact that it is rational in him to believe P. Whether a believing is a correct believing is an objective, not a subjective, matter. (Thomson 2008, p.116)

Believing that *p* is thus supposed to be correct just in case *p*, as its being the case that *p* is the condition under which the belief that *p* is deserved. Moreover, according to Thomson, a similar line of thought reveals that other mental states have correctness conditions too. For example, *trustworthy people* are deserving of trust, so trustworthy people figure in the correctness condition of trust; *admirable objects and people* are deserving of admiration, so admirable objects and people figure in the correctness condition of admiration; *desirable objects and states of affairs* are deserving of desire, so desirable objects and states of affairs figure in the correctness condition of desire; and so on.

With the idea of the correctness conditions of various mental states in place, Thomson proceeds to ask what it is for there to be *reasons for* being in some mental state. Her answer to this question comes in two parts. In the first part, she insists that we can give an independent characterization of what it is for something to be a reason for believing. Reasons for believing are, she tells us, considerations that provide "evidence for", or that "make probable", or that "lend weight to" the truth of the propositions believed (Thomson 2008, p.130). As I understand it, Thomson's suggestion here is in the same spirit as the earlier suggestion that we can analyze the concept of a reason for believing in terms drawn from the theory of perception and the theory of epistemic

rationality. The crucial point, in other words, isn't her specific proposal about what these terms are, but rather her acknowledgement that we can analyze the concept of a reason for belief in other terms. Equipped with an analysis of what reasons for believing are, the second part of Thomson's answer comes in the form of a quite general analysis of reasons for being in a mental state that has a correctness condition. A consideration is a reason for being in a mental state that has a correctness condition, she tells us, just in case that consideration is evidence for, or makes probable, or lends weight to, the truth of the proposition that is that mental state's correctness condition (Thomson 2008, p.131).

How plausible is Thomson's quite general analysis of reasons for being in mental states with correctness conditions? Assuming, for the sake of argument, that she gets the specifics of the analysis of reasons for belief right, it certainly does seem to be plausible. For, on that assumption, reasons for belief are considerations that lend weight to the truth of the proposition believed, as the truth of the proposition believed is the correctness condition of belief; reasons for trust are considerations that lend weight to the truth of the proposition that the person trusted is trustworthy, as the truth of the proposition that the person trusted is trustworthy is the correctness condition of trust; reasons for admiration are considerations that lend weight to the truth of the proposition that the object of admiration is admirable, as the truth of the proposition that the admired object is admirable is the correctness condition of admiration; and reasons for desire are considerations that lend weight to the truth of the proposition that the desired object is desirable, as the truth of the proposition that the object of desire is desirable is the correctness condition of desire.

Note that we can now better understand what was going on in the case of reasons for action. We saw earlier that Scanlon thinks that reasons for action reduce to reasons for desire, where the desires in question might or might not be desires to act in ways that one could act by acting on those very desires: in other words, the desires might or might not be, or be components of, intentions. But now that we have the quite general analysis of reasons before us, it turns out that reasons for action can all be seen to reduce, at an even more fundamental level, to reasons for believing that acting in the relevant way is desirable. In most cases, we are able to act in the way that is desirable by acting on the desire that there is a reason for having, and in these cases, we can say that reasons for action reduce not just to reasons for believing that acting in the relevant way is desirable, but that they also reduce to corresponding reasons for intentions. The reason to write a paper about *On What Matters* is a case in point. But in other cases, though our reasons for action reduce to reasons for believing that acting in the relevant way is desirable, and though there are reasons to desire that we act in the relevant way, these reasons are not reasons for having corresponding intentions, because there are no corresponding intentions. The reason to spontaneously squawk like a chicken is a case in point.

Let's take stock. What's especially important about Thomson's quite general analysis of reasons for being in mental states with correctness conditions is that it tells us that and why:

(General Thesis) All reasons-for are reasons for believing (Thomson 2008, p.130) is true, and hence why reasons for belief, trust, admiration, desire, aims, actions, and so on, are all of a kind. They are all of a kind because reasons for belief reduce in the way

we suggested at the outset, and reasons for mental states with correctness conditions other than belief are reasons in virtue of satisfying the very same condition that reasons for belief satisfy: that is, they are reasons in virtue of being considerations that support the truth of the proposition that is the relevant mental state's correctness condition. What's especially important about Thomson's quite general analysis of reasons for being in mental states with correctness conditions is thus that it tells us that and why the concept of a reason isn't a disjunction of something fundamental and something non-fundamental: it tells us that and why the concept isn't a ragbag. This is a major advance.

4. *The analysis of value*

So far I have been talking as though Thomson's suggestion that there are reasons not just for beliefs, but also for trust, admiration, and desire, is completely unproblematic. I have done this because I have been assuming both that she is right that these mental states have correctness conditions, and that her general analysis of what it is for there to be reasons for mental states with correctness conditions is along the right lines, as it provides us with the guarantee we need that the concept of a reason isn't a ragbag. There is, however, a residual puzzle.

Why do trust, admiration, and desire, have, as their correctness conditions, the trustworthiness of the person trusted, the admirability of the admired object, and the desirability of the desired object? Without an answer to these questions, we don't yet fully understand why there are reasons for trust, admiration, and desire. These questions become especially urgent when we realize that certain answers we might plausibly give, in at least some of these cases, don't readily extend to other cases. For example, the trustworthiness of the person trusted, and the admirability of the admired object, might be correctness conditions of trust and admiration because trusting someone and admiring something are, *inter alia*, a matter of believing that the person trusted is trustworthy, and believing that the admired object is admirable. But though this would certainly suffice to explain why trust and admiration have the correctness conditions that they do, the explanation proffered would not extend to the case of non-instrumental desire. It would not extend because it is manifestly implausible to suppose that non-instrumentally desiring something is, *inter alia*, a matter of believing that the desired object is non-instrumentally desirable. Quite in general, for reasons Hume appreciated all too well (§1), non-instrumental desires do not have beliefs as parts. Beliefs and non-instrumental desires are distinct existences, as for any non-instrumental desire and belief pair, there is a possible world in which someone has the one, but lacks the other, and *vice-versa* (Smith 1987, Smith 1988).

So let's confront what seems to me to be the core of the puzzle head-on. Why does desire have, as its correctness condition, the desirability of the desired object? Is there some Objectivist theory of desirability that provides us with the needed answer? Thomson herself thinks that goodness is an *attributive*: the kinds of thing that are judged to be good determine the standards for things of those kinds, and hence to judge things of a kind 'good', she thinks, is just to say that they meet those standards. A good toaster is one that evenly warms and browns bread without burning it; a good burglar is one who enters buildings from which he is prohibited entry and leaves again without getting caught; a good move in chess is one that efficiently contributes towards checkmating, or failing that, drawing with, one's opponent; and so on. Attributive goodness is thus an

Objectivist conception. Might the concept of desirability just be the concept of attributive goodness? Would this explain why desire has the desirability of the desired object as its correctness condition?

The answer is that it would not. Consider the kind: *next move for a burglar to make in certain circumstances*. There might well be considerations that lend weight to the truth of the proposition that a good next move for a burglar to make in certain circumstances would be to (say) lie to the security guard. But this doesn't entail that there are reasons for the burglar to desire to do that, or reasons for us to desire that we do that if we find ourselves in the burglar's circumstances, or reasons for us to desire that the burglar does that, or indeed reasons for anyone else to desire anything either. There seems to be no connection at all between reasons for believing that something is attributively good and reasons to have desires concerning those things. The concept of desirability in play when we suppose that desire has the desirability of the desired object as its correctness condition thus isn't the concept of attributive goodness.

Is there some other version of Objectivism that provides the needed answer? In *On What Matters*, Parfit offers his own Objectivist theory of goodness. He tells us that when we call something

good, in what we can call the *reason-implying* sense, we mean roughly that there are certain kinds of fact about this thing's nature, or properties, that would in certain situations give us or others strong reasons to respond to this thing in some positive way, such as wanting, choosing, using, producing, or preserving this thing. (Parfit 2011, p.38)

Parfit's idea is that the property of something's being good, in the reason-implying sense, is the second-order property of there being reasons in the fundamental sense for having certain desires concerning that thing, and for performing certain actions, actions like choosing, using, producing, or preserving the thing. Might the concept of desirability just be the concept of goodness in the reason-implying sense? Would this explain why desire has the desirability of the desired object as its correctness condition?

To begin, note that Parfit's analysis of goodness in the reason-implying sense is defective even by his lights, as there is built-in redundancy. We have already seen that Parfit seems not to think that there are any reasons in the fundamental sense for performing actions. He thinks that reasons for action reduce to reasons for desires or aims. Having already said that goodness is a matter of having fundamental reasons for having certain desires, he therefore adds nothing to his analysis of goodness in the reason-implying sense by mentioning reasons to act. Any reasons for action are already implicit in the mention of reasons for desires. Worse still, though, at least for present purposes, the analysis also gets things the wrong way around. Following Thomson, we are now supposing that reasons for desire reduce to reasons for believing desirable. But this means that the concept desirability cannot in turn reduce to the concept of what there is reason to desire, otherwise believing that something is desirable would be a matter of believing that there is a reason to believe that there is a reason to believe that...and so on and so forth *ad infinitum*. The content of the belief would never get fixed. The concept of desirability in play is thus not the concept of goodness in the reason-implying sense either.

Might we suppose that the concept of desirability is itself a primitive property, not analyzable at all, not even in terms of the concept of a reason? This was, in effect, Moore's own preferred version of Objectivism (Moore 1903). But Moore's view famously makes it completely obscure why reasons to believe that something is good, in his sense, should be reasons to desire the things that have that property. After all, as we have already seen, there are certain goodness properties, namely, attributive goodness properties, for which there is clearly no such connection. So why should the primitive property that Moore posits have such a connection? Why isn't it, in this respect, just like attributive goodness, completely divorced from anything to do with what there are reasons to desire?

We could canvas other Objectivist possibilities, but the obvious alternative at this point is to revert to some sort of Subjectivism and see whether it fares any better. What we need is an analysis of desirability that at no point presupposes facts about what there is reason to desire, but an analysis that none the less makes it transparent why desire has as its correctness condition the desirability of the desired object. The version of Subjectivism that I have argued for elsewhere naturally suggests itself. According to that version of Subjectivism, when we believe that it is desirable that p , what we believe is (very roughly) that we would desire that p if we were fully informed and rational (Smith 1994). Our desiring that p thus has the desirability of p as its correctness condition for the simple reason that this is what it would be for our desiring that p to match the desire of our fully informed and rational counterparts. In other words, incorrect desires are ill-informed or irrational.

To definitively spell out this version of Subjectivism, we must of course say what it is for someone to be fully rational, and, in so doing, we mustn't presuppose facts about what there is reason to desire. Again, though, as I have argued elsewhere, we might well be able to do this (Smith 2011). So far we have in effect been assuming that being a rational is a matter of having and exercising a pair of capacities: the capacity to believe for reasons and the capacity to realize desires. What's remarkable, though, is that these capacities don't fully cohere with each other. If one has certain sorts of desires, then the exercise of the capacity to realize desires can lead one to fail to exercise the capacity to believe for reasons—think of cases of wishful thinking—and *vice versa*. So if we suppose that our being *fully* rational is a matter of our having a *maximally* coherent psychology, then it seems that that might in turn be possible only if we have certain non-instrumental desires, non-instrumental desires whose realization ensures that, when we realize our desires, their realization never comes at the cost of our failing to exercise our capacity to believe for reasons. Such non-instrumental desires would themselves be, as it were, *coherence-makers*.

Here is not the place to rehearse all of the arguments required to make this suggestion fully convincing. Suffice it to say that considerations like those just adduced seem to me to tell in favour of our supposing that, in order to have a maximally coherent psychology, one has to have a dominant non-instrumental desire that one does not interfere with one's present or future exercise of one's capacities to believe for reasons or realize desires, or anyone else's for that matter, and one also has to have a dominant non-instrumental desire that one does what one can to ensure that one has the capacities to believe for reasons and realize desires, both in the present and in the future, to exercise,

and others too. These non-instrumental desires are necessary because only so can the two capacities be brought into coherence with each other. If this line of argument, or something in the vicinity of it, is successful, and if, as seems plausible, the experience of agony, whether in the present or the future, is guaranteed to undermine one's exercise of the capacity to believe for reasons and realize desires—this, we might now suppose, is what the difference between mere pain and agony consists in—then it would follow that everyone who has a maximally informed and coherent psychology would have to have an instrumental desire to avoid agony.

The idea that Subjectivism supports some such argument amounts to the idea that certain desires, and in particular the desire to avoid agony, whether in the present or the future, can be derived from non-instrumental desires that are partially *constitutive* of what it is to be maximally rational, constitutive because their presence in a psychology makes for the coherence of that psychology. But if the desire to avoid present and future agony can be derived from non-instrumental desires that are partially constitutive of what it is to be maximally rational, then it follows that everyone would desire to avoid agony if they were fully rational, and it follows from that, together with the analysis of desirability already proposed, that everyone is such that, if they were to believe that it is desirable to avoid agony, then what they believe would be true. Reasons for believing that avoiding agony is desirable—reasons of the kind just spelled out—would therefore turn out to be reasons to desire to avoid agony. Subjectivism would in this way enable us to reduce reasons for desires to reasons for belief in just the way Thomson proposes. But what's key here, to repeat, is the fact that the analysis of desirability at no point presupposes facts about what there is reason to desire, while none the less making it transparent why desire has as its correctness condition the desirability of the desired object.

When Parfit objects to Subjectivism, he ignores the possibility that some version of the theory might support an argument of the kind just proposed. Here is his initial statement of the sort of Subjectivism about reasons for action that I myself might endorse.

Some other Subjectivists appeal, not to what would best fulfill or achieve our desires or aims, but to the choices or decisions that we would make after carefully considering the facts. These people also make claims about how it would be rational for us to make such decisions. According to what we can call

(C) what we have most reason to do, or decisive reasons to do, is the same as what, if we were fully informed and rational, we would choose to do.

But this claim is ambiguous. Subjectivists and Objectivists may both claim that, when we are trying to make some important decision, we ought to deliberate in certain ways. We ought to try to imagine fully the important effects of our different possible acts, to avoid wishful thinking, to assess probabilities correctly, and to follow certain other procedural rules. If we deliberate in these ways, we are *procedurally* rational.

Objectivists make further claims about the desires and aims that we would have, and the choices that we would make, if we were also *substantively* rational. These

claims are *substantive* in the sense that they not about *how* we make our choices, but about *what* we choose. There are various telic desires and aims, Objectivists believe, that we all have strong and often decisive object-given reasons to have. To be fully substantively rational, we must respond to these reasons by having these desires and aims, and trying to fulfill or achieve them if we can. Deliberative Subjectivists make no such claims. These people deny that we have such object-given reasons, and they appeal to claims that are only about procedural rationality. (Parfit 2011, pp.62-63)

And here is his attempt to respond to an argument of the kind I've just outlined.

Deliberative Subjectivists appeal to what we would want and choose after some process of informed and *rational* deliberation. These people might argue:

(A) We all have reasons to have those desires that would be had by anyone who was fully rational.

(B) Anyone who was fully rational would want to avoid all future agony.

Therefore

We all have a reason to want to avoid all future agony.

As I have said, however, such claims are ambiguous. Objectivists could accept (B), because these people make claims about *substantive* rationality. According to objective theories, we all have decisive reasons to have certain desires, and to be substantively rational we must have these desires. These reasons are given by the intrinsic features of what we might want, or might want to avoid. We have such a decisive object-given reason to want to avoid all future agony. If we did not have this desire, we would not be fully substantively rational, because we would be failing to respond to this reason.

Subjectivists cannot, however, make such claims. On subjective theories, we have no such object-given reasons, not even reasons to want to avoid future agony. Deliberative Subjectivists appeal to what we would want after deliberation that was *merely procedurally* rational. On these theories, *if* we have certain telic desires or aims, we may be rationally required to want, and to do, what would best fulfil or achieve these desires or aims. But, except perhaps for the few desires without which we could not even be agents, there are no telic desires or aims that we are rationally required to have. We can be procedurally rational whatever else we care about, or want to achieve. (Parfit 2011, p.78)

What is remarkable about this passage is that Parfit doesn't distinguish, as he should, between the basic claims Subjectivists make in stating their theory, on the one hand, and the claims that they make because they think that they can be derived from their basic claims, on the other.

True enough, a Subjectivist might initially state his theory in largely procedural, or better perhaps *structural*, terms. The concept of coherence as applied to a psychology is, let's agree, a structural notion. The basic claims of the kind of Subjectivism I myself embrace, to repeat, are that a fully rational psychology is one in which the capacities to believe for reasons and realize desires are both possessed to the maximal degree, and are fully exercised, and are brought into coherence with each other. But having made these basic claims, the Subjectivist might argue in the way proposed that, for these capacities to be brought into coherence with each other, we would have to have a non-instrumental desire that we do certain things. I have suggested that we would have to have a non-instrumental desire that we do not interfere with our own, or anyone else's, exercise of the capacities to believe for reasons or realize desires, and that we would also have to have a non-instrumental desire that we do what we can to ensure that we and others have such capacities to exercise. Given that future agony would frustrate such non-instrumental desires, we would have to instrumentally desire that we avoid future agony.

The Subjectivist's argument for the conclusion that we all have a reason to desire to avoid agony would then proceed as follows. The fact that we would all desire that that we avoid future agony if we were fully informed and rational entails that avoiding future agony is desirable; and it follows from this that there are reasons to believe that future agony is desirable, and that these reasons are reasons to desire that we avoid future agony; and it follows from this, via Parfit's own analysis of reasons for action in terms of reasons for desires or aims, that we have reasons to avoid future agony. Though Parfit is therefore right that the claim that we all have a reason for desiring that we avoid future agony is not one of the basic claims that the Subjectivist makes, he is wrong that the Subjectivist must deny this claim. On the contrary, the Subjectivist might end up making this claim precisely because he derives it from his basic claims. This kind of Subjectivism seems to me much more promising than Parfit's own Reasons Fundamentalism.

Conclusion

Let me summarize the conclusions argued for in this paper. Parfit thinks that the concept of a reason for belief is fundamental, but there is good reason to believe that this is a mistake, and that the concept of a reason for belief can be analyzed. Hume's reductive theory provides the model. Parfit also thinks that the concept of a reason for desire is fundamental, but if the concept of a reason for belief can be analyzed, then this too turns out to be a mistake. To suppose otherwise is to embrace an implausible conception of reasons as a ragbag, a disjunction of something fundamental and something non-fundamental. Parfit thinks that the concept of goodness can be analyzed in terms of the concept of reasons in the fundamental sense for desiring and acting, but this turns out to be a mistake even by his own lights, as he doesn't really think that reasons for acting are fundamental, and it is also a mistake for the reasons we have already given, namely, because there are no fundamental reasons for desiring. Parfit thinks that Subjectivists have to deny the claim that everyone has a reason to desire to avoid future agony, but this is a mistake, as all that Subjectivists have to deny is that this is one of the basic claims they make in stating their theory. They can quite happily agree that this claim is derivable from the basic claims they make, and some Subjectivists provide arguments for this conclusion. Finally and most importantly, Parfit dramatically underestimates the

appeal and power of Subjectivism. The appeal and power of Subjectivism lies in the fact that it alone promises to explain something that desperately needs explaining, namely, why the correctness condition of desire is the desirability of the desired object.

REFERENCES

Hume, David 1740: *A Treatise of Human Nature* (Oxford: Clarendon Press, 1968)

Moore, G.E. 1903: *Principia Ethica* (Cambridge: Cambridge University Press)

Parfit, Derek 2011: *On What Matters*, Volume One (Oxford: Oxford University Press).

Scanlon, Thomas M. 1998: *What We Owe To Each Other* (Cambridge: Harvard University Press)

Smith, Michael 1987: "The Humean Theory of Motivation" in *Mind*, 96, pp.36-61

_____ 1988: "On Humeans, Anti-Humeans and Motivation: A Reply to Pettit" in *Mind*, 97, pp.589-595.

_____ 1994: *The Moral Problem* (Oxford: Wiley-Blackwell, 1994)

_____ 2011: "Deontological Moral Obligations and Non-Welfarist Agent-Relative Values" in *Ratio*, XXIV, pp.351-363

Thomson, Judith Jarvis 2008: *Normativity* (Open Court Publishing)