ON *NORMATIVITY*

MICHAEL SMITH

Judith Jarvis Thomson's *Normativity* (2008) is a formidable book in terms of both content and style (all otherwise unattributed page references in what follows are to this book). The structure is modular, as though to encourage the reader who wants to dip in and out. But as you read it becomes clear that dipping in and out would not be a good idea. Questions that occur to you are sometimes addressed immediately, but sometimes the answers can only be inferred from what's said elsewhere. Moreover, sometimes the questions themselves only occur to you after you've reached the end and tried to figure out how the whole story fits together. Nor should this be surprising. Thomson's ambition is nothing less than to explain normative thought and talk in all its diversity and complexity. Her account includes disparate elements that relate to each other in sophisticated ways. My aim in what follows is to prompt her to say more about how some specific elements fit together.

Let's start with Thomson's objection to Consequentialism, as that leads naturally into many of her more positive claims. Consequentialism is the view that we can understand the directives that are true of agents in terms of the values of the possible worlds in which those agents perform the actions that are available to them as options. If there is a uniquely best possible world in which the agent pursues one of his options, then, according to Consequentialism, he ought to perform the action he performs in that possible world; if various possible worlds are tied for having most value, then he ought to perform the action that he performs in one of those possible worlds—the disjunction is obligatory, but each disjunct is permissible; and if the values of the possible worlds are incommensurable, or on a par (Chang 2002), then more complex directives will be true of the agent: perhaps he will face a dilemma, or a situation which we can describe, but for which we have no directive term in English.

What is important about this statement of Consequentialism, at least for Thomson's purposes, is the assumption that possible worlds have either the property of being good to some degree (p.12), where their possession of this property explains where they sit in a ranking of all of the possible worlds from best to worst, or the comparative property of being better or worse than certain other worlds, where their possession of this property explains their place in an overall ranking in terms of value (p.62). This is important because Thomson thinks that that assumption is false: there is no property of being a good possible world, and nor is there a property of one possible world's being better than another. Consequentialism is thus, quite literally, incoherent.

Thomson's argument for this conclusion comes in three parts. In the first part she argues that:

> (i) Being a good K is being good *qua* K …
>
> (ii) There is such a property as being good *qua* K if and only if K is a goodness-fixing kind …
>
> (Strong Conclusion) There is such a property as being a good K if and only if K is a goodness-fixing kind.
>
> (pp.19-21)

Thomson argues for (i) by the elimination of alternatives. Something's being a good K_1 can't just amount to its being good *and* a K_1 because, if it did, it would follow from something's being a good K_1 and its being a K_2 that it is a good K_2, whereas that does not follow: a good tennis player who plays chess need not be a good chess player (pp.3-6). The obvious diagnosis, here, is that judgements of goodness are made relative to the kind in question: a good K is good *qua* K. Thomson then argues for (ii) on the grounds that, when we unpack the idea of something's being good *qua* K, what we discover is that "being a K … [must] itself [set] the standards that a K has to meet if it is to be good *qua* K" (p.21). This is what it is for K to be a goodness-fixing kind. Being a chess player, for example, itself sets the standards for being a good chess player, so being a chess player is a goodness-fixing kind. In the second part of her argument, Thomson argues, on analogous grounds, that "there is such a relation as 'being a better K' just in case K is a goodness-fixing kind" (p.60). And in the third and final part she argues that not all kinds are goodness-fixing kinds. Pebble, for example, is not a goodness-fixing kind (p.21), nor is smudge, or cloud, or shade of grey, or piece of wood (p.22), and nor are event, or act, or fact, or state of affairs, or, crucially, possible world (p. 25). As she puts it, as regards possible worlds:

> I know of no account of what a possible world is such that what being a possible world is itself sets the standards that a possible world has to meet if it is to be good *qua* possible world. Isn't anything that consists in being a way the world might be, or being a suitably large state of affairs, as good a specimen of that kind as any other?
>
> (p.26)

But if there is no such thing as being a good possible world, or of being a better possible world than some other, then, Thomson concludes, Consequentialism is in deep trouble.

The Consequentialist might, of course, insist that possible world *is* a goodness-fixing kind. But that response doesn't seem very promising. Thomson does tell us that there are various ways in which a kind could be goodness-fixing, but, at least on plausible assumptions, the kind possible world isn't a goodness-fixing kind in *any* of these ways. For example, one way is for the kind in question to be a *functional* kind (p.20). Another is for the kind in question to have been *designed for a certain purpose* (p.20). Yet another is for the kind in question to be a *biological* kind (p.20). And yet another—and this turns out to be the most important, for Thomson's broader purposes—is for the kind in question to *have capacities that members of the kind might or might not exercise* (pp.20-21). Human beings, for example, aren't just members of a biological kind, but are also beings with the capacity to be just and generous—that is to say, for short, they have moral capacities. Morally good human beings, Thomson tells us, are therefore those that exercise their moral capacities (p.79) But with this list of the ways in which a kind could be a goodness-fixing kind before us, Thomson thinks it is clear that the kind possible world is not a goodness-fixing kind, for possible world is not a functional kind; it is not a biological kind; and possible worlds do not possess moral capacities that they might or might not exercise. Of course, if reality were God's creation, then perhaps possible worlds could have been designed for a purpose. But Consequentialism itself is supposed to be neutral on that particular metaphysical claim.

If Consequentialism is not to be abandoned, then Consequentialists need to come up with some response to Thomson's objection. But what might their response be?

Myself I think that they should insist that there can be such a thing as a good possible world even though possible world is not a goodness-fixing kind. In order to see how they might argue for this conclusion, it is worth remembering that Thomson's argument was initially aimed at G. E. Moore. Moore explicitly combined Consequentialism with the view that being good is a simple non-natural property of states of affairs (or possible worlds) (pp.2-3, pp.10-17), a view of the nature of being good that is not only at odds with the account that Thomson offers, but also at odds with accounts of what it is for something to be good that are given by others. So let's ask whether Thomson's argument work equally well against other versions of Consequentialism, versions that aren't committed to Moore's particular view about what it is for something to be good (we will return to Moore's view presently). For example, does it work equally well against R. M. Hare's version of consequentialism? If not, why not? What is it about Hare's view that enables him to make sense of something's being a good K without K's being a goodness-fixing kind?

The answer is that Thomson's argument does not work against Hare's version of consequentialism for reasons that Hare explains in his response to a similar argument put forward by Geach in "Good and Evil" (1957):

> [Geach thinks that] 'good' has the same descriptive meaning in the expressions 'good knife' and 'good stomach' although, as he and I agree, 'the traits for which a thing is called "good" are different according to the kind of thing in question' (p.37). He thinks that this can be so because, although there are no common traits, the meaning of the word 'good', taken in conjunction with that of the word 'knife' or that of the word 'stomach', enables us to specify the traits which things of these kinds have to have in order to be called 'good.' …There is a certain class of words (called in [*The Language of Morals*] 'functional words') for which this manoeuvre is very inviting. 'A word is a functional word if, in order to explain its meaning fully, we have to say what the object it refers to is *for*, or what it is supposed to do'. Examples of functional words are 'auger', 'knife' and 'hygrometer'…Where 'good' precedes a functional word, most of what Geach says is correct. He passes uncritically, however, from the truth about functional words to the much more sweeping claim (which is unjustified) that the same can be said of all uses of 'good'… 'Good' often precedes words which are not functional. In such cases, in order to know what traits the thing in question would have to have in order to be called good, it is not sufficient to know the meaning of the word. We have also to know what standard is to be adopted for judging the goodness of this sort of thing; and this standard is not even partly (as in the case of functional words) revealed to us by the meaning of the word which follows 'good'.
>
> (Hare 1957/1972, pp.33-34)

Since Thomson is right that the standards for judging possible worlds is not revealed to us merely by reflection on the kind of thing that a possible world is, it follows that when Consequentialism says that certain possible worlds are better than others their use of 'better' must mean something different from what it means when it is used in conjunction with some goodness-fixing kind term. Of course, this obliges Consequentialists to say what the difference in meaning consists in, and in giving that explanation, they must not fall into the trap of saying that it is a version of what Thomson calls "good-modified". But, given his non-cognitivism, it is clear what Hare, at any rate, thinks that Consequentialists should say to explain the difference, and it is also clear why, if they do say this, they will not fall into that trap.

Certain uses of 'good', Hare would say, like the one used in the statement of Consequentialism, accord with the doctrine known as *Judgement Internalism*. The scope

and strength of this doctrine are controversial, but in Hare's view it is a doctrine about the judgements that are in play when we deliberate about what to do. What Judgement Internalism says is that, first, such judgements are about the values of the possible worlds that are available to agents through the pursuit of their options, and second, that when someone judges that a certain available possible world is better than alternatives, he has a preference for that possible world over the alternatives (Hare 1981 pp.20-24). Note, however, that if Judgement Internalism, so understood, is indeed a constraint on the legitimate use of 'good' as it is used when we deliberate about what to do, then this suggests one way in which we might mark off the sense of 'good' that is in play in the statement of Consequentialism from the sense of 'good' that Thomson is concerned to analyze.

Imagine someone who judges that a particular K is good *qua* K. It plainly does not follow from this that he prefers that particular K to alternatives. A good virus is presumably one that replicates in a whole variety of hosts. However it does not follow that someone who judges a particular virus to be a good one—imagine a scientist who discovers that some virus he is studying in the lab replicates in nearly every host—prefers that virus to alternatives. He might consistently despise that virus, precisely because it is so good *qua* virus. Judgement Internalism is thus not a constraint on the meaning of good *qua* K. Nor is it a constraint on the meaning of good-modified. Someone who judges that a certain brand of toothpaste would (say) be good for use in infecting whole populations with a certain virus need have no preference for that brand of toothpaste over other brands. But if the sense of 'good' as it is used in the statement of Consequentialism is different from the senses that Thomson is concerned to analyze—if the former is constrained by Judgement Internalism, whereas the latter are not—then her objection to Consequentialism lapses.

Can we say something more positive about the meaning of 'good', so used? Hare thinks we can. For note that if we combine Judgement Internalism with a view of moral judgement as expressive of belief pure and simple, then we fly in the face of Hume's idea that belief and desire are distinct existences: that is, we fly in the face of the idea that no matter what beliefs we imagine an agent to have, and what desires, we can always imagine him having those beliefs with quite different desires, and vice versa. Since Hare thinks that we should not fly in the face of Hume's idea, he concludes that the best explanation of Judgement Internalism is that evaluative judgements themselves are not just expressions of beliefs about features possessed by the possible worlds that are judged better, but that they also expressions of a preference for those possible worlds over alternatives, preferences that those who have the same beliefs as they have when they make their evaluative judgements may have or lack. In other words, Hare thinks that when we focus on these deliberative uses of 'good', we learn that judgements involving such uses have to be given a non-cognitivist—what Thomson calls an 'Expressivist'—analysis. This is the more positive account of the meaning of 'good' to which we are led, according to Hare.

Thomson in effect anticipates this line of reply, but she does not address it in the terms just described. Rather, she tells us that

> …it cannot be too strongly stressed that "good" does not mean something different in moral and nonmoral linguistic contexts. The adjective "good" is not ambiguous. It means the same in "good

> government" as it does in "good umbrella". (Just as the word "big" means the same in "big camel" and "big mouse".) It means the same in "morally good plan" as it does in "strategically good plan". "Morally good plan" means something different from "strategically good plan", of course, but that is not because "good" means something different in those two expressions; the difference is entirely due to the difference in what modifies "good" in them.
>
> (p.37)

Moreover, she thinks that any Expressivist worth his salt would agree with her. A view like the one just described in connection with Hare, which insists on a non-cognitivist analysis of deliberative uses of 'good', but which allows that talk of good knives, good augers, good hygrometers, and good umbrellas may be explained along the lines Thomson suggests, is, she thinks, insufficiently "meaty".

> [I]f [the expressivist] thinks that all is well with the property of being a good umbrella, then what is distinctive about his Expressivism? There isn't much meat in it.
>
> There *is* a meaty Expressivism. Its friends think, not merely that there is no such property as goodness, but also that there is no such property as being a good umbrella.
>
> (p.38)

Having identified this meaty version of Expressivism, she then goes on to argue, convincingly, that it is implausible (pp.38-58). It is implausible because there is no reason to suppose that *whenever* people call things 'good', they have a preference for things of that kind. Remember again what we said earlier about good viruses.

The objection to Expressivism is thus that no credible rationale can be given for what Thomson calls:

> (Containment Thesis) 'Believing' that X is a good K is a complex that contains a want to do something.
>
> (p.53)

According to Thomson, the Containment Thesis is the thesis to which meaty Expressivists are committed. But I cannot see why Expressivists should take on the burden of defending the Containment Thesis. The Containment Thesis is similar to Judgement Internalism, but in fact the two doctrines are worlds apart. The Containment Thesis posits a connection between judgements of being a good K *qua* K and desires, whereas Judgement Internalism posits a connection between judgements of goodness as 'good' is used in deliberative contexts and desires. Of course, Thomson thinks that there is no such separate use of 'good'. We might call this her 'No Ambiguity Thesis'. But, at this stage of the argument, she is supposed to be arguing for that conclusion, not assuming it as a premise. The upshot is that the less meaty version of Expressivism Thomson ignores is left intact. Her objection to Expressivism thus lapses too.

Let's now reconsider Thomson's objection to Moore's theory. The question is whether the detour via Expressivism shows us how a Moorean might reply. Moore was, of course, no Expressivist. He would presumably have thought that someone who judges certain possible worlds to be better than others expresses a belief about the comparative distribution of the simple non-natural property of goodness within those possible worlds. Thomson's objection to this idea, you will recall, is that there is no such property because possible world is not a goodness-fixing kind. But we can now see how Moore might

have replied, and this reply is in essence the line taken by modern intuitionists. Moore might have replied that there is a distinctive feature of the use of 'good' in deliberative contexts which is best explained by the supposition that talk of 'good' possible worlds doesn't presuppose that possible world is a goodness-fixing kind, but which presupposes instead that it picks out a simple non-natural property of possible worlds. Moreover, the discussion of Expressivism suggests an obvious candidate for what this feature might be: Judgement Internalism.

I said earlier that when we combine Judgement Internalism with the view that moral judgement is an expression of belief, we fly in the face of Hume's idea that belief and desire are distinct existences. Modern intuitionists wonder what's supposed to be wrong with that. John McDowell, for example, thinks that we should argue in the other direction.

> [In]...urging behaviour one takes to be morally required, one finds oneself saying things like this: 'You don't know what it means that someone is shy and sensitive.' Conveying what a circumstance means in this loaded sense, is getting someone to see it in the special way in which a virtuous person would see it. In the attempt to do so, one exploits contrivances similar to those one exploits in other areas where the task is to back up the injunction 'See it like this': helpful juxtapositions of cases, descriptions with carefully chosen terms and carefully placed emphasis, and the like... No such contrivances can be guaranteed success, in the sense that failure would show irrationality on the part of the audience. That, together with the importance of rhetorical skills to their successful deployment, sets them apart from the sorts of thing we typically regard as paradigms of argument. But these seem insufficient grounds for concluding that they are appeals to passion as opposed to reason: for concluding that 'See it like this' is really a covert invitation to feel, quite over and above one's view of the facts, a desire which will combine with one's belief to recommend acting in the appropriate way.
>
> (McDowell 1978, pp.21-2)

McDowell admits that

> Failure to see what a circumstance means, in the loaded sense, is of course, compatible with competence, by all ordinary tests, with the language used to describe the circumstance; that brings out how loaded the notion of meaning involved in the protest is.
>
> (McDowell 1978, p.22)

and he therefore concludes that the ordinary tests we have for individuating agents' ways of thinking about their circumstances are inadequate.

> To preserve the distinction we should say that the relevant conceptions are not so much as possessed except by those whose wills are influenced appropriately.
>
> (McDowell 1978, p.23).

In other words, seeing things in the distinctive way in which a virtuous person sees them—believing what the virtuous person believes about the circumstances of action she faces—*entails* having certain desires about how things turn out in those circumstances (see also McDowell 1979). This is in effect a version of Judgement Internalism that posits an internal connection between an agent's beliefs about the various specific modes of value that he is confronted with, modes of value that may only be cognizable by a virtuous person, and his will.

What McDowell insists is so special about these beliefs is their content: more specifically, what's special are the evaluative features of the situations about which the virtuous person has knowledge, for an agent's believing that there are situations with such evaluative features entails his having corresponding desires. In supposing that there are such evaluative features, McDowell thus flies in the face of Hume's idea that belief and desire are distinct existences. But, as I said, he can see no reason why we should worry about that when reflection on the sorts of evaluative beliefs that the virtuous person has should convince us that Hume's idea is mistaken: the beliefs and desires of the virtuous person *are not* distinct existences. To the extent that McDowell holds a contemporary version of Moore's view—like Moore, McDowell insists that evaluative features are irreducible: that is, they cannot themselves be identified with some complex set of natural features—it therefore follows that Thomson's objection to Mooreanism lapses as well. For all that she says, there is a simple non-natural property of possible worlds which is such that, when you believe, or perhaps know, that it is instantiated, it follows that you desire that that possible world be actual. I cannot see how anything Thomson says bears on whether this is so.

In the light of this discussion, it should be clear how a variety of other theorists should react to Thomson's analysis of 'good' as well. Consider, to give just one example, Thomas M. Scanlon's view that something is good just in case it has the higher-order property of having some property that provides a reason to desire it, or to admire it, or to have some other favourable attitude towards it (Scanlon 1998 pp.95-97). On the assumption that there is such a higher-order property, and on the further assumption that someone can believe that something is a good K without thinking that there are reasons to desire it, or to admire it, or to have some other favourable attitude towards it—think again about a scientist who is confronted with a good virus—Scanlon also seems to have strong grounds for denying the No Ambiguity Thesis. Indeed, Scanlon's analysis seems tailor-made to capture what's important about the meaning of 'good' as it is used in deliberative contexts. This is because deliberation looks to be a matter of figuring out what reasons there are for desiring things, or for preferring certain things over others, and then coming to desire those things, or to prefer them over others, in the light of those reasons. Moreover, since desires and preferences can have whole ways things might be (ie possible worlds) as their objects, Scanlon's analysis also seems tailor-made to make sense of Consequentialism's basic idea that goodness is a property of possible worlds (see also Smith 2009).

Indeed, it is worth remarking that Scanlon's analysis commits him to a weaker and more plausible version of Judgement Internalism than that discussed thus far, a version that Thomson also seems to be committed to denying in so far as she holds the No Ambiguity Thesis. According to this weaker version of the doctrine, if an agent judges that something is good, in the sense relevant to deliberation, then he desires that thing *insofar as he is rational*. This weaker version of the doctrine is more plausible than the version discussed thus far because it doesn't fly in the face of Hume's view that belief and desire are distinct existences. Someone evidently could have the belief that something is good without desiring it, so the belief and the desire are distinct existences, just as Hume insists. It is just that he would be, to that extent, irrational. Scanlon's analysis commits him to the truth of this weaker version of Judgement Internalism because, on the

plausible assumption that a rational agent is someone whose desires are sensitive to his beliefs about such reasons as he has for desiring, and that an irrational agent is one whose desires are insensitive to his beliefs about his reasons for desiring, he is committed to thinking that when agents believe that there are reasons for desiring something, and hence that that thing is good, then they do indeed desire that thing in so far as they are rational. Thomson, by contrast, if she sticks with the No Ambiguity Thesis, is committed to denying even this weaker version of Judgement Internalism. There need, after all, be no irrationality in an agent's failing to desire what he believes to be good in Thomson's sense. Think again about the scientist who judges that a particular virus is a good virus, or that a particular brand of toothpaste would be good for use in infecting whole populations with a certain virus.

With Scanlon's view fresh in our minds, let's now turn to some rather different claims Thomson argues for in *Normativity*. As I said, Scanlon holds that something is good just in case it has the higher-order property of having some property that provides a reason for (let's say) desiring it. Never mind that Thomson denies that that's what it is for something to be good. Focus on a different question. Does Thomson think that that higher-order property exists? Somewhat surprisingly, the answer is that she does. Indeed, Thomson thinks that there are reasons for having all sorts of mental states and she provides us with an account of what it is for there to be such reasons. She talks through the case of reasons for belief and reasons for trust in some detail. In what follows, however, in the interest of laying out the similarities and differences between her views and Scanlon's, I will additionally focus on what this commits her to saying about reasons for desiring.

Thomson's starting point is an evaluation of a different kind from those discussed thus far, namely, the standard of correctness for believing: or, more colloquially, the standard by which we judge what it is right to believe. She tells us that:

> 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.
>
> (p.116)

Nor, according to Thomson, is believing the only kind of mental state that can be correct or incorrect. The same goes for many other kinds of mental state, though not for all. Trusting, admiring, and desiring can all be correct or incorrect because, as with believing, there is something that counts as rightly trusting, rightly admiring, and rightly desiring. Feeling dizzy, by contrast, cannot be correct or incorrect. Nothing counts as rightly feeling dizzy (p.132). So what makes it the case that certain mental states can be correct or incorrect, whereas others can't?

According to Thomson, whether or not a mental state of a certain kind can be correct or incorrect turns on the nature of the mental state itself, as the kind itself determines what it is for an object to be *deserving* of that mental state (p.116). Just as the nature of belief itself tells us that it is *truths* that deserve belief; so the nature of trust tells us that *trustworthy people* are deserving of trust; the nature of admiration itself tells us

that *admirable objects and people* are deserving of admiration; and the nature of desire itself tells us that *desirable objects and states of affairs* are deserving of desire. The relevant contrast here is with mental states like feeling dizzy, as we have no idea what to make of the suggestion that dizziness is deserved. Thomson calls mental state kinds that can be correct or incorrect correctness-fixing kinds (pp.130-132). However, as she is quick to point out, this is quite different from saying that they are goodness-fixing kinds. A believing, for example, is not worse *qua* believing if what's believed is false; nor is a trusting of someone untrustworthy worse *qua* trusting; nor is an admiring of someone who isn't admirable worse *qua* admiring; and nor is a desiring of something that isn't desirable worse *qua* desiring (pp.111-112). All of this just goes to show how rich and diverse the evaluative domain really is. Correctness and goodness are quite different evaluative features.

With the idea of a correctness condition for mental states in place, we are now in a position to explain what, according to Thomson, it is for there to be *reasons for* being in some mental state. Thomson's core idea is that:

> (General Thesis) All reasons-for are reasons for believing
>
> (p.130)

In the case of reasons for believing, she thinks, we can give an informative account of what such reasons are. Reasons for believing are considerations that provide "evidence for", or that "make probable", or that "lend weight to" the truth of the propositions believed (p.130). More generally, if more abstractly, Thomson thinks that:

> For X to be a reason for just anyone to V_{mind} is for X to lend weight to ψ_{Vmind}
>
> where ψ_{Vmind} is the proposition such that for a V_{mind}-ing to be a correct V_{mind}-ing is for ψ_{Vmind} to be true.
>
> (p.131)

What this tells us, in the case of belief, is that for p to be a reason for just anyone to believe that q is for p to lend weight to q, because q is the proposition that has to be true for a believing that q to be correct. But it also tells us what reasons for other mental states are. In the case of trust, for example, for p to be a reason for just anyone to trust so-and-so is for p to lend weight to the proposition that so-and-so is trustworthy, because the proposition that so-and-so is trustworthy is the proposition that has to be true for a trusting so-and-so to be correct. In the case of admiration, for p to be a reason for just anyone to admire such-and-such is for p to lend weight to the proposition that such-and-such is admirable, because the proposition that such-and-such is admirable is the proposition that has to be true for an admiring such-and-such to be correct. And in the case of desire, for p to be a reason for just anyone to desire that q is for p to lend weight to the proposition that it is desirable that q, because the proposition that it is desirable that q is the proposition that has to be true for a desiring that q to be correct.

Let's suppose we accept Thomson's analysis of what it is for there to be reasons for various mental states. What does the analysis show and what doesn't it show? One thing it does show is that, whenever p is a reason for some non-belief mental state, it is also a reason for believing something in particular, namely, that the proposition that is the

correctness condition of that mental state is true. However note that it doesn't show that there are no reasons for non-belief mental states—indeed, an assumption of the analysis is that there are such reasons—and nor does it show that, though there are reasons for non-belief mental states, reasons for non-belief mental states reduce to reasons for believing. *All* it shows is that reasons-for always come in pairs, where one of the pairs is a reason for believing something in particular. Now I suspect that Thomson thinks that this undersells her analysis. More specifically, I suspect she thinks that her analysis does show more, as it shows that reasons for belief are explanatorily prior to reasons for non-belief mental states; this, I take it, is the intended import of the General Thesis. But it is important to see that this conclusion has not been established either.

The reason why emerges when we contrast what Thomson has to say with Scanlon's famously pessimistic view that we cannot say anything interesting about what it is for a consideration to be a reason. Thomson sums up Scanlon's view as follows:

> So on Scanlon's view, there is no such thing as an interesting analysis of what it is for X to be reason for ϕ, whatever ϕ may be… No interesting analysis, that is. Scanlon thinks that for X to be a reason for ϕ is for X to be a consideration that counts in favour of ϕ, but that if you want to know how a reason counts in favour of what it is a reason for, the only answer available is the not at all helpful: "a reason counts in favour of what it is a reason for by providing a reason for it."
>
> (p.127)

As Scanlon sees things, reasons for believing and reasons for other mental states are thus on a par: both are just considerations that count in favour. Against this, Thomson does succeed in showing that Scanlon is too pessimistic, as she manages to tell us something far more interesting about how reasons-for count in favour. Reasons for believing count in favour by providing evidence for, or making probable, or adding weight to the proposition believed, and reasons for other mental states count in favour in such a way that they are always paired with a reason for believing that those mental states' correctness conditions obtain. What's true is thus that, if no more independent account of how reasons for non-belief mental states count in favour could be given, then it would follow that our grip on how reasons for non-belief mental states count in favour would come entirely via their relationship to reasons for believing. This would establish a kind of explanatory priority of reasons for believing. But since Thomson doesn't ever ask whether some independent account can be given of how reasons for non-belief mental states count in favour of them, much less establish that no such independent account can be given, the explanatory priority of reasons for belief remains an open question.

At this point, two strategies are available to those who wish to insist that reasons for believing are not explanatorily prior to reasons for non-belief mental states. One is to give the needed independent account of how reasons for non-belief mental states count in favour. Can this be done? My own view is that, at least in the case of reasons for intrinsically desiring, it can. The intrinsic desires for which there are reasons, it seems to me, are all and only those intrinsic desires that are constitutive of being ideally rational, and the considerations that are the reasons for those intrinsic desires are fixed by those desires' contents. Thus, for example, if having an intrinsic desire for pleasure is constitutive of being ideally rational, then it seems to me that it follows that the nature of pleasure is the reason for intrinsically desiring pleasure. The nature of pleasure is a consideration that counts in favour of intrinsically desiring pleasure because it is related

in the right way to the content of a desire that is constitutive of being fully rational. Though Thomson is thus right that another idea is explanatorily prior to the idea of a reason for desiring, the idea that is explanatorily prior is the idea of being ideally rational, not the idea of a reason for believing (see Smith 2010).

The other strategy is to deny that Thomson has succeeded in giving an independent account of what it is for something to be a reason for believing. In so far as we have any grip at all on what it is for a consideration to be evidence for, or to make probable, or to lend weight to a proposition, the opponent might insist that we have that grip only by way of our grip on the idea of the consideration in question's counting in favour of *believing* that proposition. The explanatory burden is being carried by the *attitude* in question, not by the ideas of being evidence for, or making probable, or adding weight to. There is therefore no explanatory asymmetry, as our grip on what it is for a consideration to count in favour of desiring is similarly fixed by the fact that it is *desiring* that the consideration is supposed to count in favour of. There is doubtless much more to say about each of these strategies. But for now it will perhaps suffice to say that, since there is clearly unfinished business here, we shouldn't agree that Thomson has established the explanatory priority of reasons for believing over reasons for other mental states. At best she has shown that reasons-for come paired in the way described.

Let me close by focusing on what seems to me to be the most remarkable feature of Thomson's discussion of reasons for mental states. What's most remarkable is that it shows that she not only agrees with Scanlon that things do possess the higher-order property of having some property that provides a reason for desiring them—or, equivalently, of having the property of being the object of a correct desiring, or of deserving being desired—but that she also agrees with him that this is an *evaluative feature* of those things. Of course, Thomson rejects Scanlon's suggestion that that evaluative feature is those things' being good, but she concedes that it is an evaluative feature nonetheless. Here is the relevant passage:

> There is room to say that an ascription of 'being a correct trusting' to Smith's trusting Alfred entails a favourable evaluative judgement, for it entails that Alfred has 'deserves being trusted by Smith'—and 'deserves being trusted by Smith' is surely a favourable evaluative property. Again, an ascription of 'being a correct admiring' to Smith's admiring Bert entails that Bert has 'deserves being admired by Smith'—and 'deserves being admired by Smith' is surely a favourable evaluative property. I have no objection to that idea.
>
> (p.119)

Similarly, then, Thomson would presumably have no objection to the idea that an ascription of 'being a correct desiring' to Smith's desiring that p entails a favourable evaluative judgement, for it entails that p has 'deserves being desired by Smith'—and 'deserves being desired by Smith' is surely a favourable evaluative property. And indeed it is a favourable evaluative property, as it is just the property of being desirable (or, perhaps better, the property of being desirable$_{\text{Smith}}$).

If this is right, however, then it seems to me that Thomson's earlier argument against Consequentialism was based on a miscommunication of ideas. That argument, you will recall, assumed that Consequentialists ascribe the property of being good to possible worlds, where 'good' in 'good possible world' means the same as 'good' in 'good K' where K is a goodness-fixing kind. This is why Thomson thought that

Consequentialism makes sense only if possible world is itself a goodness-fixing kind, which it isn't. Against this, I suggested that Consequentialists might insist that their theory should be stated in terms of a use of 'good' as 'good' is used in deliberative contexts, and I pointed out that Scanlon's suggestion that something is good just in case it has the higher-order property of having some property that provides a reason to desire it seems the perfect candidate for what's picked out by 'good' in such contexts. But it now turns out that this line of reply is one that Thomson herself could have given on the Consequentialist's behalf. For she believes that the higher-order property that Scanlon identifies with the property of being good exists, and she agrees that it is an evaluative property, she just doesn't think that it is the property of *being good*: instead, she thinks that it is the property of *being desirable*. Thomson might therefore have restated Consequentialism in more plausible terms before attacking it.

So stated, Consequentialism is the view that that we can analyze the directives that are true of agents in terms of the *desirability* of the possible worlds in which those agents perform the actions that are available to them. As far as I can tell, Thomson is in no position to think that Consequentialism, so understood, is incoherent. Indeed, it seems that her own account of what it is for a possible world to be desirable tells us not just that the view is coherent, but also what its attractions are. According to Consequentialism, so understood, someone's ϕ-ing is impermissible just in case, when you compare the possible world in which he ϕs with the possible worlds in which he performs each of the alternative acts available to him, what you discover is that there is a possible world in which he performs an alternative act that has a feature that makes it deserve to be desired more than the possible world in which he ϕs. The question to which the Consequentialists quite reasonably demand an answer is what could possibly justify the agent's ϕ-ing in these circumstances. How could it be permissible for an agent to act in a certain way if his acting in that way did not deserve being desired?

Just to be clear, I am not suggesting that no answer can be given to this question. I am simply insisting that the Consequentialists are right to demand that we have an answer to it. Moreover, as I understand it, Thomson's own theory of directives is supposed, in effect, to provide an answer. She tells us that:

> (HB-V_{act} Thesis) If A is a human being, then for it to be the case that A ought to V_{act} is for it to be the case that if A knows at the time what will probably happen if he V_{act}s and what will probably happen if he does not, then he is a defective human being if he does not.
>
> (p.216)

The idea behind (HB-V_{act} Thesis) is that, since human beings have moral capacities like the capacity to be just and generous, it follows that human being is both a goodness-fixing kind and a directive kind (pp.207-218). It is a goodness-fixing kind because the kind human being itself sets the standard for what it is to be a better or worse human being: a human being is better or worse depending on how just and generous he is. And it is a directive kind because the kind human being itself tells us what it is to be a defective human being: to be a defective human being is to lack the virtues of justice or generosity (p.218). So, leaving out some detail that isn't strictly relevant to the point at hand, the actions that agents ought to perform are all and only those which are such that, if they fail to perform them, then they are unjust or ungenerous. This, I take it, would be

Thomson's answer to the question to which the Consequentialists demand an answer. Her answer would be that it doesn't follow from the mere fact that an agent performs an action that doesn't deserve being desired that he is unjust or ungenerous. Or, to put the point another way, though it certainly does follow from the fact that an agent desires to perform actions that don't deserve to be desired that he isn't as good a person-who-possesses-the-capacity-to-desire-what-deserves-desiring as a good such person can be, it doesn't follow that he is defective.

There is a great deal to say about this, but for now I would like to focus on just one main point, a point which has just emerged. I would have thought that, once we admit that certain desirings are correct desirings, then we would have no alternative but to explain what the moral capacities of human beings are in those terms. Thomson may, of course, be right that, in so far as we have moral capacities, we have capacities for justice and generosity. But if she is right, then that will have to be because the objects of the desires of those who are just and generous deserve desiring. The order of explanation will have to go from a person-who-possesses-the-capacity-to-desire-what-deserves-desiring—in Thomson's terms, this is the "super-kind" to which we belong, so the directives analyzed in its terms will defeat those analyzed in terms of its sub-kinds (pp.213-214)—to human being with moral capacities, to human being who is capable of justice and generosity. (I note, in passing, that this is equivalent to insisting that we explain the virtues in terms of the idea of what there is reason to desire, not vice versa.) Assuming that this is so, it follows that the idea of a defective human being, which Thomson claims to get straight out of the idea of a human being as a possessor of moral capacities, must also somehow be derived from the idea of a person-who-possesses-the-capacity-to-desire-what-deserves-desiring. But I don't see how we could derive the idea of a defective human from the idea of a person-who-possesses-the-capacity-to-desire-what-deserves-desiring. I am therefore skeptical that the idea of a defective human being can be invoked to do the work that Thomson needs it to do in her analysis of directives.

Of course, this leaves us without an answer to the question to which the Consequentialists demand an answer. How could it be permissible for an agent to act in a certain way when his acting in that way does not deserve being desired? To be perfectly honest, I don't know how to answer that question. One possibility, however, is that the answer is best given in terms of an ambiguity in our concept of permissibility. Sometimes when we say that an agent's ϕ-ing is permissible we are talking about the directives that are true of him, and perhaps the Consequentialist is right that what this means is that the possible world in which the agent ϕ's deserves being desired, though we leave it open that possible worlds in which he takes other options available to him also deserve being desired. But sometimes when we say that an agent's ϕ-ing is permissible we aren't really talking about the directives that are true of the agent, but are rather talking about the directives that are true of *other* agents (Darwall 2006; Bedke forthcoming). What we mean is that others ought to leave the agent free to ϕ if he chooses, even if his so choosing means that he desires to do something that doesn't deserve desiring. In the Consequentialist's terms, this would amount to the claim that the possible world in which others leave him free to ϕ deserves being desired.

To repeat, whether a story along these lines could be worked up into a full-blown answer to the question to which the Consequentialists demand an answer, I do not know.

But what I do know is that, unlike Thomson's answer, which trades on the suspect idea of a defective human being, an answer along these lines would not trade on that idea. Moreover, such an answer would have the additional advantage of having been constructed out of what seems to me to have emerged from Thomson's story as the most explanatorily basic element, namely, the idea of a person-who-possesses-the-capacity-to-desire-what-deserves-desiring.

REFERENCES

Bedke, Matt forthcoming: "Passing the Deontic Buck", *Oxford Studies in Metaethics* edited by Russ Shafer-Landau (Oxford: Oxford University Press).

Chang, Ruth 2002: "The Possibility of Parity" *Ethics* (112) pp.659-88

Darwall Stephen, 2006: *The Second-Person Standpoint: Morality, Respect, and Accountability* (Cambridge: Harvard University Press).

Geach, Peter 1957: "Good and Evil" *Analysis* (17) 1957 pp.33-42.

Hare, R. M. 1957/1972: "Geach: Good and Evil" originally in *Analysis* (17) 1957 pp. 103-111, reprinted in his *Essays on the Moral Concepts* (London: Macmillan; University of California Press, 1972).

Hare, R.M. 1981: *Moral Thinking* (Oxford, Oxford University Press)

McDowell, John 1978: "Are Moral Requirements Hypothetical Imperatives?", *Proceedings of the Aristotelian Society* Supplementary Volume (52) pp.13-29.

McDowell, John 1979: "Virtue and Reason", *The Monist* (62) pp.331-50.

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

Smith, Michael 2009: "Two Kinds of Consequentialism" in *Philosophical Issues* (19) pp.257-272.

_____ 2010: "Beyond the Error Theory" in *A World Without Values: Essays on John Mackie's Moral Error Theory* edited by Richard Joyce and Simon Kirchin (New York: Springer).

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