both the Program, and All Souls, for providing ideal research environments. I am also grateful to numerous audiences for their reactions to this paper's ancestors, including those at Brown University, Harvard University, New York University, the University of Massachusetts at Amherst, St. Andrews University in Scotland, the Program in Economics, Justice and Society at the University of California at Davis, the Society for Ethics and Legal Philosophy in Cambridge, and the World Health Organization. My memory is too faulty to properly acknowledge all those who have given me useful comments on this topic, but they include Baruch Brody, Gerald Cohen, Tyler Cowen, Roger Crisp, Keith DeRose, Fred Feldman, Charles Fried, David Gauthier, James Griffin, Susan Hurley, Shelly Kagan, F.M. Kamm, Liam Murphy, Thomas Nagel, Ingmar Persson, John Roemer, Tim Scanlon, Amartya Sen, Ernest Sosa, Peter Unger, and Peter Vallentyne. John Broome, Stuart Rachels, and J. David Velleman deserve special mention for their influence on this essay. Finally, Derek Parfit initially inspired my work on this topic, and he has consistently been both my biggest source of encouragement and my most penetrating critic.

2 More generally, any relation R will be transitive, if and only if, for any a, b, and c, if aRb and bRc, then aRc. 'Taller than' is a standard example of a transitive relation, since if Andrea is taller than Becky, and Becky is taller than Claire, then Andrea is taller than Claire. By contrast, 'being the birth mother of' is clearly not a transitive relation, since Andrea's being the birth mother of Becky, and Becky's being the birth mother of Claire, does not entail – and in fact is incompatible with – Andrea's being the birth mother of Claire.

3 This example assumes that the value of $2 is not significantly different from the value of $1. If we imagine a scenario where the difference of one dollar meant the difference between life and death, then this wouldn't be an example of case II.

4 For the sake of discussion, I assume, contrary to fact, that there is no significant diminishing marginal utility of income between $1,000,000 per year and $2,000,000 per year.

5 Satisficers are people who believe there is a point where 'enough is enough,' after which they eschew a maximizing strategy in decision making. For an interesting discussion of the attractions of satisficing, see Michael Slote's *Beyond Optimizing: A Study of Rational Choice* (Harvard University Press, 1989).

6 Vallentyne's example is contained in his illuminating review of John Broome's *Weighing Goods* (Basil Blackwell Inc., 1991); see 'The Connection Between Prudential and Moral Goodness', *The Journal of Social Philosophy* 24 (1993): 105-28.

7 The argument of this section is taken from my 'Weighing Goods: Some Questions and Comments', *Philosophy and Public Affairs* 23, no. 4 (1994): 350-380. However, most of the discussion of the argument is new.

8 Or at least as not less valuable, which is all one really needs to defend this position.

9 *Reasons and Persons* (Oxford University Press, 1984), p. 75.

10 *Reasons and Persons*, p. 388.

11 I have been working on such a book for many years now, tentatively entitled *Rethinking the Good, Moral Ideals, and the Nature of Practical Reasoning*. Unfortunately, the ground is treacherous, and so far my work has yielded more problems than solutions.

# The Resentment Argument[1]

Michael Smith

The holy grail of moral philosophy must surely be an argument that would show how and why certain facts, when properly appreciated by people, no matter who those people are or what their antecedent inclinations might happen to be, rationally require of those people a certain kind of response. Perhaps the required response would be a desire that those facts be realized, or perhaps indifference to those facts being realized, or perhaps an aversion to those facts being realized. That doesn't matter. What matters is not the nature of the response rationalized, but rather the fact that appreciation of the specified facts would rationalize some response or other.

The argument I have in mind would, of course, straddle the divide between meta-ethics and normative ethics. Like arguments in meta-ethics it would proceed without making any, or at any rate without making any undefended, normative assumptions. But like arguments in normative ethics it would enable us to draw a substantive normative conclusion. In other words, it would be no less than an argument taking us from an 'is' to an 'ought'. While no one much likes to describe themselves as attempting to derive an 'ought' from an 'is' these days, my own view is that at least some of the best work that has been done in moral philosophy in the last thirty years or so is best understood in just these terms. My aim in this essay is to focus attention on one such argument, the role reversal argument, which proceeds by way of asking the question 'How would you like it if someone did that to you?'.

The role reversal argument seems an especially fitting topic for a volume of essays in honor of Ingmar Persson. I first got to know Ingmar as the author of a critique of a well known presentation of this argument, a presentation that was subsequently described quite explicitly as an attempt to derive an 'ought' from an 'is' (Robinson 1982). Persson's target was, of course, Hare's famous attempt to derive utilitarianism from the formal features of moral thinking: that is, from the fact that, if thinking is to count as moral thinking at all, then the thinker must be trying to figure out what she can overridingly and universally prescribe (Hare 1981). Whenever I teach Hare's argument to my students, I teach it alongside Persson's careful and devastating critique (Persson 1983). It seemed to me when I first read Persson's critique, and it still seems to me now, that it provides an

excellent illustration of just how much we can learn when we subject an apparently convincing argument – an 'intuitive' argument, as we might say – to rigorous analysis.

The argument is a fitting choice for another reason as well. For while Hare's argument, partially under Persson's influence, is no longer seriously supposed by anyone to have any hope whatsoever of allowing us to derive an 'ought' from an 'is', another presentation of the role reversal argument from much the same period is still alive and well. The alternative presentation of the argument first appeared in Chapter Nine of Thomas Nagel's *The Possibility of Altruism*. Here is what Nagel says.

> The rational altruism which I shall defend can be intuitively represented by the familiar argument, 'How would you like it if someone did that to you?' It is an argument to which we are all in some degree susceptible; but how it works, how it can be persuasive, is a matter of controversy. We may assume that the situation in which it is offered is one in which you would not like it if another person did to you what you are doing to someone else (the formula can be changed depending on the type of case; it can probably be used, if it works at all, to persuade people to help others as well as to avoid hurting them). But what follows from this? If no one is doing it to you, how can your conduct be influenced by the hypothetical admission that if someone were, you would not like it?
>
> It could be that you are afraid that your present behavior will have the result that someone will do the same to you ... It could be that the thought of yourself in a position similar to that of your victim is so vivid and unpleasant that you find it distasteful to go on persecuting the wretch. But ... why cannot such considerations motivate you to increase your security against retaliation, or take a tranquilizer to quell your pity, rather than to desist from your persecutions?
>
> There is something else to the argument; it does not appeal solely to the passions, but is a genuine argument whose conclusion is also a judgement. The essential fact is that you would not only *dislike* it if someone else treated you in that way; you would resent it. That is, you would think that your plight gave the other person a reason to terminate or modify his contribution to it, and that in failing to do so he was acting contrary to reasons which were plainly available to him. In other words, the argument appeals to a *judgement* that you would make in the hypothetical case, a judgement applying a general principle which is relevant to the present case as well. It is a question not of compassion but of simply connecting, in order to see what one's attitudes commit one to. (Nagel 1970, pp.82–3)

Let's call this the 'resentment argument'.

According to Nagel, the resentment argument provides merely 'intuitive' support for rational altruism. But though in the remainder of *The Possibility of Altruism* he goes on to give a much more complicated and sophisticated argument for that same conclusion, it is the intuitive argument that has had a lasting impact. Nagel himself has since abandoned the sophisticated argument (Nagel 1986), for example, but he quite happily continues to repeat the intuitive argument (Nagel 1987). More recently, Christine Korsgaard relies exclusively on the intuitive argument at a crucial point in *The Sources of Normativity* (Korsgaard 1996). In response to her own question 'Now how do we get from here' (where 'here' is an argumentative situation in which we are a long way short of rational altruism) 'to moral obligation?' (which is an argumentative situation in which we have derived some obligations towards others, and have hence committed ourselves to some form of rational altruism), Korsgaard responds 'This is where Thomas Nagel's argument ... comes into its own' (Korsgaard 1996, p.142). The argument she goes on to give is none other than the resentment argument.



Despite its evident impact it is, I think, less than clear both what the resentment argument is meant to be and whether, when properly understood, it has any real force. Accordingly, my aim in this paper is to spell out the resentment argument in some detail and to provide an evaluation of it. Though I would be delighted if subjecting the resentment argument, Nagel's version of the role reversal argument, to rigorous analysis enabled us to learn as much as we learned from Persson's analysis of Hare's version of that argument, I will be content if the conclusion is the more plonking one that Nagel's intuitive argument for rational altruism, much like Hare's for utilitarianism, collapses under close scrutiny.

## Clarification of the resentment argument's main premise and conclusion

Let's begin by clarifying the main premise of the resentment argument.

I am to imagine that I have acted in some way such that I would not like it if someone acted in that way towards me. Nagel suggests that it might be a situation in which I have harmed another person, or a situation in which I have failed to provide the other with some benefit. To make things simple, in what follows I will focus on cases of harming another. Nagel does not say so, but it must also presumably be imagined that the alternative action was available to me in the circumstances. In other words, I am to imagine a situation in which I harm another when I could have failed to harm him instead.

The crucial observation to make about this main premise of the resentment argument is that, being about a harm done in circumstances in which a harm might not have been done, it is a premise that falls fairly and squarely on the 'is' side of the 'is-ought' gap. This premise is about a non-evaluative matter of fact. Acceptance of it does not, all by itself, entail acceptance of the rightness or wrongness of what was done, and nor does it fix a subject's orientation to what was done either. Those who believe themselves to be harming someone in the circumstances described could be in favor of their so acting, or they could be

indifferent to so acting, or they could be averse to so acting. Acceptance of the premise does not yet tell us which of these states they are in, still less does it tell us whether their being in one or another of these states would be rationally required or forbidden.

Consider now the conclusion of the resentment argument. Nagel tells us that the argument is meant to be 'persuasive'. In context, it is plain what he means by this. He means that the argument is meant to have a rational influence on the conduct of those who appreciate its force. The conclusion of the resentment argument is thus supposed to be a motivation to act. Moreover, and much more importantly for the argument's claim to be an *argument*, Nagel tells us that when the resentment argument exerts this influence on conduct, it exerts it by way of supporting a 'judgement', or, if you like, by way of supporting an intermediate propositional conclusion. The intermediate conclusion is the claim that, when I harm another in circumstances in which I could have failed to harm them, I act in a way that I have a reason not to act and hence have a reason to stop so acting.

Just as we saw that the main premise of the resentment argument fell on the 'is' side of the 'is'–'ought' gap, we can therefore see that the intermediate conclusion of the argument, which is a claim about what there is a reason to do, a claim whose acceptance is in turn supposed to rationalize the main conclusion of the argument, a corresponding motivation, evidently falls on the 'ought' side. So if the resentment argument contains no further premises – in other words, if everything is meant to follow a priori from the main premise – then it seems that the argument does indeed purport to take us from a non-evaluative premise to an evaluative conclusion. The resentment argument thus really does look like it is meant to be the holy grail of moral philosophy: little wonder that Korsgaard thought she should appeal to it at that crucial juncture in *The Sources of Normativity*!

Even if the resentment argument is everything it purports to be, however, note that this is not to say that those who accept the 'is' main premise of the resentment argument cannot fail to accept the 'ought' conclusion. Since the argument purports to be an *argument*, they may of course fail to accept that conclusion. They may fail to accept it either by failing to accept the intermediate conclusion that they have a reason not to do what they did, or, if they accept that intermediate conclusion, by failing to be correspondingly motivated. The important point is simply that, on the assumption that the resentment argument is everything it purports to be, to the extent that people who accept the 'is' premise fail to accept the 'ought' conclusion they thereby become liable to rational criticism.



## Moving beyond the main premise: role reversal

Consider now the first move beyond the main premise of the resentment argument.

Informally, remember, the argument proceeds by making us confront the counterfactual question 'How would you like it if someone did that to you?' Put slightly more formally, what this counterfactual question assumes is that it follows from the fact that I have harmed another person in circumstances in which I might not have harmed them, that there is a possible situation which is exactly like this one in various respects, including the nature of the harm that is done, but which differs in that I am the one who is harmed and another person is the one who does the harming.

Note that this is an 'is'–'is' move, not yet a move from an 'is' to an 'ought'. More precisely, it is a move from an 'is' that characterizes actuality to an 'is' that characterizes a mere possibility. Even so, it might well be questioned whether the move is valid. Suppose, for example, that I am a doctor and that what I have done is harm a woman by, say, damaging her reproductive system during an operation. Can we really suppose that there is a possible situation in which I suffer that harm? That would seem to require, falsely, that there is a possible situation in which my natural reproductive organs are those of a woman. Or suppose that I harm an Australian Aboriginal by making a racist remark about the genetic pool from which he came. Can we really suppose that there is a possible situation in which I suffer that harm? That would once again seem to require something impossible, namely, that there is a possible situation in which I come from the genetic pool of the Australian Aboriginals.

Having noted this problem, however, it seems to me that we can safely put it to one side. Either there is a relevant sense of 'possibility' in which these claims do state genuine possibilities, notwithstanding the fact that in the most familiar sense of 'possibility' they do not, or we can ascend to a more general characterization of the harm done so that it is plausible to suppose that, in the more familiar sense of 'possibility', it is possible for me to suffer a harm so characterized. Of course, once we ascend to this more general characterization of the harm done it might be extraordinarily difficult to specify the *respect* of similarity in non-evaluative language. The easiest way to characterize the similarity might well be in evaluative terms, for example by saying that the harm done to me in the possible case is just as bad as the harm done to the person I harm in actuality. But even if this is right it seems to me that we should resist drawing the conclusion that there is no *non*-evaluative respect of similarity in such cases. Indeed, given that a quite radical particularism is false, it seems to me that there must be some non-evaluative respect of similarity (Jackson, Pettit and Smith 2000). It is the possibility of these sorts of harms that we must countenance, harms which must be identical in non-evaluative respects, even if it is difficult for us to

characterize the nature of such harms in English without recourse to evaluative language.

To sum up: we should grant the first move beyond the main premise of the resentment argument. It does indeed seem to follow from the fact that I have harmed another person in circumstances in which I might not have harmed them, that there is a possible situation which is exactly like this one in respect of the harm that is done – or, at least, similar in crucial non-evaluative respects – but which differs in that I am the one who is harmed and another person is the one who does the harming.

## Moving beyond the main premise: feelings of resentment

The second move beyond the main premise is the observation that, in the possible situation imagined in which it is me who is harmed, and another person who does the harming, I do not just find the fact that I have been harmed by the other unpleasant, or a cause of feelings of insecurity, or to be something that I dislike. Rather I find that I resent what he does to me.

The second move thus presents itself as another 'is'–'is' move. The fact that someone feels resentment is a non-evaluative fact about them, albeit, as we will see, a fact which presupposes that the person who feels resentment is committed to certain evaluative claims. I will have more to say about this in the next section. Even putting the connection between resentment and evaluation to one side, however, the transition is still problematic. For we must ask why we should suppose that I would have any such feelings at all in the possible situation in which it is me who is harmed. What is the connection supposed to be between imagining myself being harmed and imagining myself feeling resentment? Various suggestions might be made.

To begin, it might be suggested that I have to imagine myself feeling resentment because, as we saw in our discussion of the first move beyond the main premise, we have to imagine a possible situation which is as similar as possible to the actual situation, except that our roles are reversed. Since we can assume that the person I harm in actuality resents my harming him, the suggestion might be, it follows that I have to imagine myself resenting him in the imagined situation in which our roles are reversed. Role reversal means not just taking on the other person's harm, but taking on his resentment as well. But I do not think that this can be right.

As I understand it, the resentment argument is supposed to show that I have a reason not to harm another person even if, for some reason, perhaps because he is so child-like in his appreciation of what happens to him, he feels no resentment at all when I harm him. If this is right, however, then it cannot be that the felt resentment that the person in the actual case feels towards me, even assuming



that that person does feel such resentment, is what explains why I have to imagine myself feeling resentment towards him when our roles are reversed. The resentment that the person in the actual case feels towards me, even assuming that he does feel such resentment, though perhaps perfectly legitimate, is therefore not something that I must take on board when our roles are reversed. The fact that he feels resentment for the harm I do to him is incidental and inessential for the resentment argument to gain purchase.

Another suggestion might be that I have to imagine myself resenting being harmed by the other because it is literally impossible for me to imagine being harmed without, thereby, imagining myself feeling resentment. But this cannot be right either. If it is so much as possible for there to be someone who feels no resentment when I harm him – (say) because he is so child-like in his appreciation of what happens to him – then there is at least one possible situation – the situation in which I am as child-like in my appreciation of what happens to me – in which I feel no resentment when he harms me. The claim that it is literally impossible for me to imagine myself being harmed without feeling resentment is therefore implausible.

A final suggestion, and I suspect that this must be what Nagel has in mind, is that I am supposed to imagine myself feeling resentment because, at the role reversal stage of the argument, I am not supposed to imagine just any old possible situation in which it is me who is harmed by another, but rather a situation which is as similar as possible to actuality in which it is me who is harmed by another. Informally, remember, the argument begins by asking us a counterfactual question: 'How would you like it if someone did that to you?' The assumption might be that since I actually feel resentment for similar harms that are done to me in actuality – after all, I am not in actuality child-like in my appreciation of what happens to me – so, in the counterfactual case in which a harm exactly like the one that I do to another is done to me, we must suppose that I would, in that case, feel resentment as well.

If this is right, however, then the resentment argument has been underdescribed. The argument must contain an extra premise. It must contain an extra premise because, without that extra premise, it is simply invalid. There is, after all, a possible world in which I accept the premise that there is a possible situation, maximally similar to the possible world in which the argument is being offered to me, a possible world in which I cause someone to suffer a harm, in which I suffer a harm exactly like that, but in which I quite correctly resist drawing the alleged conclusion that, in that possible situation, I would feel resentment. There is such a possible world because, in some such possible worlds in which the argument is being offered to me, completely child-like as I am in my appreciation of what happens to me. I don't ever feel resentment and hence it simply isn't true that I would feel resentment. The truth of the premise thus doesn't entail the truth of the conclusion all by itself.

What the extra premise needs to capture is the crucial assumption that the resentment argument is only being offered to people who, not being child-like in their appreciation of what happens to them, do in fact feel resentment when they experience harms like the one that I have caused the other person. But what, exactly, would this extra premise be? One possibility is that the extra premise is simply a statement to the effect that none of the possible explanations of why someone might fail to experience resentment obtain. In other words, the extra premise would state that (for example) I am not child-like in my appreciation of what happens to me, and, if there are other conditions whose obtaining would prevent me from feeling resentment, then it would state that none of these conditions obtain either. Another possibility is that the extra premise is simply a more bald statement to the effect that I am a person who feels resentment when I experience a harm like the one that I have caused the other person.

Whichever of these is the ultimate form of the relevant premise, the crucial point to make here is that some such extra premise is required. To be sure, it is another 'is' premise, not an 'ought' premise. The fact that I feel resentment is a non-evaluative fact about me, albeit, as we will shortly see, a fact which presupposes that I am committed to certain evaluative claims. As will then eventually become clear, the fact that some such extra premise is required is thus problematic.

## Moving beyond the main premise: evaluation

The third move beyond the main premise of the resentment argument is this. Granting that I feel resentment in the possible situation in which I am harmed by another, it is supposed to follow that I negatively evaluate his harming me, that is, that I judge his doing so to be undesirable.

Note that this is once again a move from an 'is' to an 'is'. The fact that I negatively evaluate someone's conduct is a non-evaluative fact about me, albeit a non-evaluative fact about my commitment to certain evaluations. Even so, we must ask why we should suppose that resentment does entail a negative evaluation. The reason, I take it, is that resentment is supposed to be a prime example of those emotions that, by their very nature, have an evaluative aspect. Other prime examples are supposed to include anger, which is said to be connected with the judgement that one has been wronged, and fear, which is said to be connected with the judgement that something is dangerous.

However, without questioning whether there are such emotions, it is important to remember that the claim that there are such emotions is subject to two quite different interpretations (contrast Solomon 1976 and Gibbard 1990). It can be interpreted as saying something quite strong, namely, that when I experience certain emotions, things don't just appear to me to be a certain way, but that I



actually judge things to be that way: I *judge* that someone has wronged me, I *judge* that something is dangerous, I *judge* that someone has acted in an undesirable way towards me, and so on. Or, alternatively, it can be interpreted as saying something much weaker: that when I am angry at someone, it *seems* to me that he has wronged me, whether or not I go on to make the judgement; that when I am afraid of something, it *seems* to me that that thing is dangerous, whether or not I go on to make the judgement; that when I resent the way that someone behaves, it *seems* to me that he has acted in an undesirable way towards me, whether or not I go on to make the judgement; and so on.

The third move beyond the main premise of the resentment argument plainly requires that we interpret the claim that emotions have an evaluative aspect in the stronger of the two ways just described. But the problem with interpreting the claim in this way is that it would seem to be quite plainly mistaken. It would seem to be mistaken because emotions that have an evaluative aspect would seem, in this respect, to bear a certain striking similarity to ordinary perceptual states. Just as when we ordinarily perceive something we systematically have appearances of objects being a certain way, so, when we experience such emotions, we systematically have appearances of value and disvalue. Crucially, though, just as when we perceive things we can know that the appearances are misleading, even while they continue to appear to us in the way that they do – think of the way in which the two lines in the Muller-Lyer illusion persist in seeming to be of different lengths even after we have measured them and convinced ourselves that they are of the same length – so, when we experience the evaluative emotions, it can continue to seem to us that there is value or disvalue even after we have convinced ourselves that such an evaluation would be mistaken. Allan Gibbard gives us a nice example of this sort of evaluative illusion.

> Most of us have experienced being angry and yet thinking that no wrong has been done, so that the anger is unjustified. In such cases, one *feels* as if a wrong had been done, but thinks that no wrong has been done. ... [I]f the anger is indeed irrational, as one thinks, then there is a belief it would be irrational to have. It would be irrational to believe that a wrong has been done. That, however, is precisely what one doesn't believe; that is why one considers one's own anger irrational. (Gibbard 1990, p.40)

If this point is agreed, however, then it follows that the resentment argument goes wrong in moving straight from the fact that I feel resentment, in the possible situation in which it is me who is harmed, to the fact that I make a negative evaluation of the conduct of the person who harms me. From the fact that I feel resentment in that possible situation all that strictly follows is that it seems to me that someone has acted in an undesirable way. But that seeming might be mere

appearance, an evaluative illusion. I might not judge that he has acted in an undesirable way.

In order to get from the fact that I would feel resentment to fact that I would judge that he acted undesirably, we therefore need to add a further premise. But what premise? Following from the earlier discussion it seems that the argument needs to make explicit the fact that the person to whom the argument is being addressed is not someone who suffers evaluative illusions in situations like this. That is why, when we ask that person to imagine how he would feel if the harm were done to him, he can not only say that he would feel resentment, but that he would judge that the person who harmed him did something undesirable.

Once again it therefore seems that we have two choices. One possibility is that the extra premise is simply a statement to the effect that the argument is being addressed to someone who does not suffer from any evaluative illusions. But since, as Gibbard points out, that amounts to the claim that the person is not irrational, it follows that the extra premise would have to be that I am not irrational. Another possibility is that the extra premise would simply be the bald statement that I judge that people who cause harms like the one done to me in the imagined case act in a way that is undesirable. This is why, in the closest possible world in which it is me who is harmed, I would judge that the harm done to me is undesirable.

This time, however, it seems that we have a decisive reason to prefer the second formulation of the extra premise to the first. For the first formulation is an 'ought' claim, whereas the second is still an 'is' claim. The fact that I make a certain evaluative judgement is a non-evaluative claim about me, albeit a non-evaluative claim about the evaluative claims to which I am committed.

## Moving beyond the main premise: reasons

Let's now consider the remaining steps in the resentment argument.

Granting both that I feel resentment in the possible situation in which I am harmed by another, and that I negatively evaluate his harming me, it is supposed to follow that, in the imagined situation, I thereby commit myself to the claim that he has a reason to stop harming me. This is plainly meant to be another 'is-is' transition: a move from a claim about a belief I would have in the imagined situation about the undesirability of the way in which the other person acts towards me to a claim about a belief I would have in the imagined situation about the reasons that that person has. From this claim, the claim that in the imagined situation I would believe that the person who harms me has a reason to stop, I am then supposed to derive, via universalization, the intermediate conclusion of the resentment argument, the conclusion that I have a reason to stop harming the person I am harming in the actual situation. Here, at last, we have a move that is



supposed to take us from an 'is' to an 'ought'. This intermediate conclusion is then, in turn, supposed to give rational support to the main conclusion of the argument: a motivation to stop.

Now it might be thought that at least some of these steps are relatively straightforward. After all, it is widely acknowledged that the term 'reason' has many senses, and that in at least one of these senses it is simply analytic that people who act in a way that is undesirable, act in a way that they have a reason not to act (see e.g. Foot 1972). If this is right, however, and if, as seems plausible, the judgement that someone acts in a way that is undesirable commits the judge to judging that it is desirable that the person stops so acting, then it follows that there is at least a sense of the term 'reason' in which the judge is committed to judging that that person has a reason to stop. If the universalization stage of the argument is acceptable then it might well be thought that we should concede these final steps in the argument. However it seems to me that it would be a grave mistake to grant all of these steps so quickly.

The first problem concerns the very first of these steps. True enough, there is a sense of the term 'reason' in which it is analytic that people who act in a way that is undesirable act in a way that they have a reason not to act. However that is not a sense of the term 'reason' that can support the main conclusion of the resentment argument which, you will remember, is a motivation to act. In order to see why this is so, consider a completely conventional mode of normative assessment, such as etiquette. Norms of etiquette are, I assume, all too often completely arbitrary from the rational point of view. Indeed, in some cases at least, they are positively pernicious, improving the position of the ruling class and undermining that of the underclass. This is why each of us can recognize, while yet quite reasonably rejecting wholesale, at least some requirements of etiquette. The mere fact that there is *a* sense of the term 'reason' in which it is analytic that what it is desirable for us to do, from the point of view of etiquette, is something that we have a reason to do, thus does nothing to show that a failure to be motivated in the way that we acknowledge ourselves to have reason to act, in *this* sense of the term 'reason', makes us liable to rational criticism. Quite the opposite. The mere existence of a reason to stop harming another, if this is the sense that the term 'reason' has, would thus likewise do nothing to give rational support to the main conclusion of the resentment argument, which is a motivation to stop.

What this shows, I think, is that the resentment argument plainly requires that we have, in the background, a particular conception of evaluative judgement, a conception that permits us to infer not just that an agent has a reason to act in a certain way from the fact that it is desirable that he acts in that way, but also that, in that sense of the term 'reason', when someone believes that they have reasons, they are liable to rational criticism if they are not correspondingly motivated.

Of course, I am happy to admit that there is such a conception of evaluative judgement, as it is a conception according to which such judgements conform to what I have elsewhere called the 'Practicality Requirement' (Smith 1994, 1997). But not everyone is happy to agree that evaluative judgements conform to such a requirement (Brink 1997). It is thus crucial to remember that the resentment argument itself presupposes that some such conception of evaluative judgement is correct.

The other much more serious problem concerns the move from 'is' to 'ought': that is, the move, via universalization, from the claim that I have to imagine myself believing that the other person has a reason to stop harming me, in the imagined situation, to the conclusion that I have a reason to stop harming the person I am harming in the actual situation. The glaring problem with this move is that, in order to be valid, it requires not just that I imagine myself *believing* that the other person has a reason to stop harming me, in the imagined situation, but that this belief is *true*. But nothing so far granted in the premises guarantees that the belief I imagine myself having is true.

Now it might be suggested that, whether or not the belief is true, I must certainly *believe* that that belief is true as I rehearse the premises of the resentment argument to myself. After all, the reason I have to imagine that I would believe that the other person has a reason to stop harming me is because this is supposed to follow from the fact that I would believe that he is acting undesirably, and the reason that I had to imagine that I would have that belief is that I in fact believe that people causing such harms act undesirably: remember how we got to make the third move beyond the main premise (section 4 above).

But while this does indeed explain why I must believe that the beliefs I imagine myself having are true, as I rehearse the premises of the resentment argument to myself, it also highlights what seems to me to be the argument's central flaw. For, as should now be plain, we can in fact by-pass all of the steps that go via role-reversal and resentment. They are all completely irrelevant. What is relevant is rather a single premise, the premise that we saw was required in order to make the third move beyond the main premise. The really crucial reasoning in the resentment argument goes like this. Premise: I judge that people who cause harms like the one that I cause to the person I harm in actuality act in a way that is undesirable. First intermediate conclusion: I should believe that what I did when I harmed that person in actuality was undesirable. Second intermediate conclusion: what I did when I harmed that person in actuality was undesirable. But this argument is plainly fallacious. We simply cannot infer from the fact that I have a belief that that belief is true. Yet that is, in effect, what the resentment argument asks us to do.



## Conclusion

Now that we have seen the premises of the resentment argument fully laid out it seems to me that we must conclude that it is a very disappointing argument indeed. What we were after, and what the resentment argument promised, is the conclusion that I have a reason to stop harming the person I am harming in actuality. But all that the premises of the resentment argument really entitle us to is the conclusion that I ought to believe that I have such a reason. Nothing in the premises supports the conclusion that this belief is true because the crucial premise that drives the conclusion is simply a premise to the effect that I believe that people who cause harms like the one that I cause to the person who I harm in actuality act undesirably. And, as with any belief, the mere fact I have this belief does nothing to show that it is true. The premises about role reversal and resentment were all completely irrelevant.

## Notes

1 An earlier version of this paper was read under the title 'In Search of the Philosopher's Stone: The Resentment Argument' at *Emotion and Value*, a conference held at Ohio State University, October 1999. I would very much like to thank all of those who participated in this splendid conference. I am especially grateful for comments received from Simon Blackburn, Miles Burnyeat, Justin D'Arms, Allan Gibbard, Philip Pettit, and Neil Tennant.

## References

Brink, David (1997), 'Moral Motivation', *Ethics*, 4–32.

Foot, Philippa (1972), 'Morality as a System of Hypothetical Imperatives' reprinted in Philippa Foot, (*Virtues and Vices*. Berkeley: University of California Press) 1978: 157–73.

Gibbard, Allan (1990), *Wise Choices, Apt Feelings* (Oxford: Clarendon Press).

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

Jackson, Frank, Philip Pettit and Michael Smith 2000, 'Ethical Particularism and Patterns' in Brad Hooker and Maggie Little (eds) *Moral Particularism* (Oxford: Oxford University Press) 79–99.

Korsgaard, Christine (1996), *The Sources of Normativity* (Cambridge: Cambridge University Press).

Nagel, Thomas (1970), *The Possibility of Altruism* (Princeton: Princeton University Press).

Nagel, Thomas (1986), *The View from Nowhere* (Oxford: Oxford University Press).

Nagel, Thomas (1987), *What Does It All Mean?: A Very Short Introduction to Philosophy* (New York: Oxford University Press).

Persson, Ingmar (1983), 'Hare on Universal Prescriptivism and Utilitarianism', *Analysis*, 43–49.
Robinson, H. M. (1982), 'Is Hare a Naturalist?', *Philosophical Review*, 73–86.
Smith, Michael (1994), *The Moral Problem* (Oxford: Basil Blackwell).
Smith, Michael (1997), 'In Defence of *The Moral Problem*: A Reply to Brink, Copp and Sayre-McCord' in *Ethics*, 84–119.
Solomon, Robert C. (1976), *The Passions* (Garden City, N.Y.: Doubleday/Anchor).

# Intrinsic Value and Individual Worth

Michael J. Zimmerman

The headline in this morning's *Daily News* proclaims, in block letters 5 cm tall:

DIANA'S DRESS CAUSES BIG SENSATION!

We all understand what it means, but should any of us take it literally?

What the headline means is that there was something about Diana's dress that caused a big sensation. The headline is intended to entice us to read further and discover what this something was. (Was the dress especially lavish? Was it shockingly revealing? Was its design outrageous? What exactly was it that caused such a stir?) Once we have learned these details, we will have a better understanding of what took place.

Suppose the dress was especially lavish, carrying a price tag of £30,000. It is this that caused the sensation. Once we have discovered this, what will our attitude be toward the claim made in the headline? Will we accept it as literally true? Will we, that is, still want to say that *the dress* caused the sensation, once we have learned that *the dress's being lavish* caused the sensation? Perhaps, but, if so, we surely wouldn't want to say that these causes are on a metaphysical par; for that would put us at risk of having to say that the sensation was causally overdetermined, which (we may assume) it was not. If we are not to be eliminativists of a certain sort and deny that the dress was, literally, a cause of the sensation, we must at least be reductionists and say that its being such a cause was nothing above and beyond some state of the dress being a cause of the sensation. Object-causation, if there is such a phenomenon at all, is metaphysically parasitic on state-causation, and so talk of the former is reducible to talk of the latter.[1]

The next morning's headline declares:

DIANA'S DRESS OF GREAT VALUE!

Is this something we should take literally?

That depends on the sort of value at issue. If it is economic value, then it seems quite natural to take it literally (and also to accept it as true; £30,000 is a lot of money for a dress). But suppose that this headline appears, not in the *Daily News*,

Persson, Ingmar (1983), 'Hare on Universal Prescriptivism and Utilitarianism', *Analysis*, 43–49.
Robinson, H. M. (1982), 'Is Hare a Naturalist?', *Philosophical Review*, 73–86.
Smith, Michael (1994), *The Moral Problem* (Oxford: Basil Blackwell).
Smith, Michael (1997), 'In Defence of *The Moral Problem*: A Reply to Brink, Copp and Sayre-McCord' in *Ethics*, 84–119.
Solomon, Robert C. (1976), *The Passions* (Garden City, N.Y.: Doubleday/Anchor).

# Intrinsic Value and Individual Worth

Michael J. Zimmerman

The headline in this morning's *Daily News* proclaims, in block letters 5 cm tall:

DIANA'S DRESS CAUSES BIG SENSATION!

We all understand what it means, but should any of us take it literally?

What the headline means is that there was something about Diana's dres that caused a big sensation. The headline is intended to entice us to read furthe and discover what this something was. (Was the dress especially lavish? Was i shockingly revealing? Was its design outrageous? What exactly was it that causec such a stir?) Once we have learned these details, we will have a better understandin of what took place.

Suppose the dress was especially lavish, carrying a price tag of £30,000. It i this that caused the sensation. Once we have discovered this, what will our attitud be toward the claim made in the headline? Will we accept it as literally true? Wil we, that is, still want to say that *the dress* caused the sensation, once we hav learned that *the dress's being lavish* caused the sensation? Perhaps, but, if so, w surely wouldn't want to say that these causes are on a metaphysical par; for tha would put us at risk of having to say that the sensation was causally overdetermined which (we may assume) it was not. If we are not to be eliminativists of a certai sort and deny that the dress was, literally, a cause of the sensation, we must a least be reductionists and say that its being such a cause was nothing above an beyond some state of the dress being a cause of the sensation. Object-causation if there is such a phenomenon at all, is metaphysically parasitic on state-causa tion, and so talk of the former is reducible to talk of the latter.[1]

The next morning's headline declares:

DIANA'S DRESS OF GREAT VALUE!

Is this something we should take literally?

That depends on the sort of value at issue. If it is economic value, then i seems quite natural to take it literally (and also to accept it as true; £30,000 is a lo of money for a dress). But suppose that this headline appears, not in the *Daily New*