

‘The Asymmetry of Good and Evil’ in Mark Timmons, ed, Oxford Studies in Normative Ethics, Vol 5, 2015, 15-37

1

The Asymmetry of Good and Evil

PHILIP PETTIT

The thesis of this paper is that there is a marked difference between doing good and doing evil—in particular, doing good and doing evil to others—as we conceive of these actions in everyday life. Doing good to others often means doing them good robustly—controlling for good effects—whereas doing evil rarely has such a robust or modal character. This asymmetry thesis has a number of implications and I will concentrate in particular on one.¹ This is that it enables us to make sense of the Knobe effect, so-called: the effect, as it appears, whereby we are generally more willing to ascribe intentionality in cases where people do evil than in otherwise similar cases where they do good.

The paper is in five sections. In the first I look at some important issues in the ontology of behavior, distinguishing between the acts or deeds whereby we produce effects and the actions or doings whereby we control for effects. In the second I show how we expect those who do good to others to control robustly for beneficial effects, being prepared at the limit to pay a cost in self-interest for securing that result. In the third I argue that we do not have the same expectation with those who do evil to others; they may aim at bringing about bad effects but not generally at any cost in self-interest. In the fourth I offer an explanation of this asymmetry. And then in the fifth section I show that the asymmetry between doing good and doing evil helps to explain the Knobe effect in a particularly satisfying way.

1. ON ACTIONS AND THEIR EFFECTS

In order to do good, in particular to do good to another, you have to perform actions that have good effects on that other, contributing to

¹ For other implications see (Pettit 2015). The argument in this paper makes use of material in that book, in particular material from chapters 5 and 6.

their welfare in some sense. And in order to do evil, in particular to do evil to another, you have to perform actions that have bad effects on that other, reducing their welfare in some sense. But if we are to understand what either involves we need to be a little clearer about our conception of actions and about the variety of effects that actions may have.

Options

Whenever you act, you choose from among a set of options—mutually exclusive and jointly exhaustive possibilities—that are accessible to you. The options are possibilities such that, depending on your preference, you are able in the circumstances on hand to realize or not realize any one of them. Suppose that you have a choice in acting between turning to the left, turning to the right, or standing still; or between sending a message by post or email or fax; or between speaking your piece or opting for silence. Each option in such a choice is something that you have the wherewithal and the knowledge to implement or enact. And in the normal case this will be manifest to you: you will not be deceived about the alternatives between which you have to choose.

On this standard conception of an option, all that determines whether or not the possibility in question is to become actual is your will or your wish; you need only try to realize it and, given how the world is, success is guaranteed. Suppose then that it is a matter of chance whether your flipping the switch will turn on the light or not; suppose that there is an element of indeterminacy about this. In that case, the option before you is that of flipping the switch, not that of turning on the light; whether or not the light comes on is an outcome associated under a certain probability with that option. If things are contingently such that flipping the switch will turn on the light, however, then turning on the light is indeed an option for you; it is something that you can just do, although you will do it only by flipping the switch.

Acts or Deeds

With this conception of an option in place, we can identify an act or a deed as the enactment or realizer of an option: an event that has the effect of actualizing the possibility represented by the option. If we think of the act as a particular event that is capable of bearing many properties, as I shall do here, then we can think of it as having a behavioral character, given by the option it realizes; an instrumental character, given by the

The Asymmetry of Good and Evil

17

way in which it realizes the option; and a variety of circumstantial effects in virtue of that behavioral and instrumental character. Under this way of thinking, then, the act with the behavioral character of turning on the light is the same act as that of flipping the switch—this gives its instrumental character—and the same act as that of surprising the burglar.²

Considered as a particular event, the act may have a variety of effects and in ascribing an effect we naturally take the act under a description that does not presuppose the effect and that highlights a property relevant to its occurrence. We do this when we describe the act as one of flipping the switch in ascribing to it the effect of turning on the light. The property of flipping the switch is causally relevant to closing the electrical circuit and offers the appropriate aspect under which to take the act in the ascription of that effect. We follow the same logic in describing the act as one of turning on the light when we ascribe the effect of surprising the burglar. It is under its aspect as an act of turning on the light, not under its aspect of an act of flipping the switch, that it surprises the burglar; after all, the burglar would not have been surprised if you had flipped the switch but the light had not come on.

The effects of an act or deed, taken as the realizer of an option, come in two broad varieties. First there are causal effects of the kind illustrated so far: the effect of the act in closing the electrical circuit, thereby putting on the light, and the effect of the act in surprising the burglar. But an act may also have effects of a non-causal or constitutive kind. This sort of effect consists in the instantiation of a distinct property by the act, where the instantiation does not come about as a result of any causal process. Thus in turning on the light you may be fulfilling a promise—say, a promise that you made to your partner to go downstairs and check on whether there is a burglar in the house. The effect whereby your act consists in keeping that promise does not materialize causally; it is logically bound to materialize in the event of your actually going downstairs to look around, given what you had said to your partner prior to descending.

Actions or Doings

All of the effects illustrated so far, causal and constitutive, are circumstantial effects of what you produce in enacting an option; they are

² In taking this line, I follow (Anscombe 1957; Davidson 1980). For an alternative way of individuating actions, see (Goldman 1970).

effects such that in the circumstances the act involved is sufficient, in virtue of its behavioral and instrumental character, to generate them. But there are other effects that you may have in acting that are not due to the character of the act as such but rather to a more complex antecedent: the combination of that act with the particular disposition out of which it was performed.

Suppose that despite having a power of interfering with me at will you exercise restraint in our interactions, allowing me to choose as I will in the normal range of personal discretion: in what we generally think of as the basic liberties that any autonomous agent ought to enjoy. Does this mean that you give me respect, as we say? Not necessarily, since any of a variety of ulterior motives may lead you to exercise restraint. It may be that you want to impress a third party, or that you are just interested in seeing how I use my discretion, when given my head. And in any such case it would not be true that you show me respect or that I enjoy respect at your hands. What I would enjoy rather is good luck or indulgence.

In order for your restraint to have the effect of showing me respect—in order for it to constitute a display of respect—you must exercise restraint out of a particular disposition. You must restrain yourself because of having renounced any act of interference in my basic liberties, putting the option of interference more or less beyond your reach. You must perform the act out of a respectful, self-denying disposition.³

Disposition-dependent effects of the kind illustrated by respect are bound to be constitutive rather than causal in character. Going downstairs to look around constitutes the keeping of a promise in the context of your earlier utterance to your partner. And exercising restraint constitutes a display of respect in the context of manifesting a disposition to renounce any form of interference with my basic liberties. The property of displaying respect attaches to the act in virtue of its origin in that disposition, not in virtue of being causally generated by the act on its own. Thus you cannot be said, strictly speaking, to give me respect by avoiding interference as you can be said to turn on the light by flipping the switch; you give me respect by avoiding interference only on the

³ That may not be enough for this effect to materialize; it may be necessary, in addition, that you are externally constrained to act in accord with the disposition, say by the law (Pettit 2015). But it is certainly necessary for giving me respect that you act out of such a disposition. The effect is disposition-dependent.

The Asymmetry of Good and Evil

19

extra assumption—perhaps salient from context—that you act out of a suitable disposition.

Given that disposition-dependent effects are not strictly effects of your acts, why should we think of them as effects that you bring about? The answer is that they are effects of what you do, not in the sense of being effects of the acts you produce, but in the sense of being effects of the control you exercise in acting. In giving me respect you actually produce an act of restraint, avoiding interference in my exercise of a basic liberty. But what is also true is that you do this because of controlling for restraint: because of bringing it about that no matter how circumstances vary in certain ways—strictly, vary according to your presumptively reliable beliefs—you will still avoid interference in my exercise of that type of choice. You might not restrain yourself if the cost were a loss of life or fortune for someone, whether that be you, me or a third party. But you would restrain yourself across a large range of scenarios where it was more or less inconvenient to avoid interference. You display restraint robustly—actually and across relevant scenarios—and in that sense control for displaying it.

Given that we describe what you produce in performing as an agent as your acts and deeds, we may speak of the controlling interventions that lead to such acts or deeds as your actions or doings. A disposition-dependent effect of giving me respect may not be an effect of your act of restraint but it certainly is an effect of your action. It is something that comes about as a result of what you control for and as a result of what you do in that richer sense.

For every action in this sense there will almost always be a corresponding act and it will share disposition-independent effects with that act.⁴ But while the action will have disposition-dependent effects essentially, those effects will be irrelevant to the act: they will not be effects of the act in the sense in which disposition-independent effects are. In performing the action of showing me respect, you produce the act

⁴ Although true in most cases, even in cases of omissions, this does not hold universally, because you may control in a virtual or standby manner for a certain effect. Think of the cowboy who rides herd on his cattle. He does nothing that helps lead his cattle in the direction of the railhead but he does control for their taking that direction. He controls for their doing so in the sense that while he does nothing in the actual case, where they go in the desired direction, he is ready to intervene should any of them wander off track. For the notion of virtual control see (Pettit 1995).

of refraining from interference. And while the action and the act will have the same disposition-independent effects, say in allowing me to choose as I wish, they will be distinguished by the fact that the respectful disposition is essential to the action but quite irrelevant to the act.

In adopting this way of thinking about human behavior, I join forces with Christine Korsgaard (2009, 12). She takes herself to follow Aristotle and Kant—and to oppose ‘many contemporary moral philosophers’—in making a distinction very like that which we have made here. She uses the term ‘act’ to pick out the option realized by the agent, as we have done, and she takes the action involved in realizing it to vary with any variation in intention or motive or policy and so with any difference in disposition-dependent effects. ‘So for instance’, she says, ‘if you choose to dance for the sheer joy of dancing, then dancing is the act, and dancing for the sheer joy of dancing is the action. We might contrast it to the different action of someone who dances in order to make money, or to dodge the bullets being shot at his feet’.⁵

What is the point of adopting this language of action or doing in addition to the language of act and deed? It reminds us that to perform an action is not just to produce an act or deed, it does not matter by what causal route. To perform an action is to produce the act or deed in manifestation of this or that intention or motive or policy. It is to produce the act in response to the consideration or set of considerations that become relevant in the presence of such a disposition. To perform the action, as we put it, is to control for the production of the act in any circumstances where the relevant considerations are registered. Thus, to show me respect is to control for the production of restraint in any circumstances, the actual ones included, where you register that my exercise of a basic liberty is at stake. Acting always involves this sort of control and the control exercised reflects the dispositions by which you are moved.⁶

⁵ I am grateful to Alec Walen for pointing out that the distinction was already made, and made in the same terms, by David Ross (2002 New Edition) in 1930.

⁶ Thinking of actions as controllers in this way may guard against a tendency in the moral philosophy of action to ignore the disposition-dependent properties of what we do and to treat questions of disposition as bearing only on the evaluation of agents, not on the evaluation of their behavior (Bennett 1995; Scanlon 2008). Barbara Herman (2011, 89–90) opposes that tendency, for example, in arguing that in ethics we should ‘make motive relevant to determinations of wrongness’.

The Asymmetry of Good and Evil

21

2. DOING GOOD

Disposition-Independent and Disposition-Dependent Benefits

There are many ways in which you may do me good. You may choose acts or deeds with a welcome behavioral or even instrumental character. And, even more saliently, you may choose acts or deeds with a variety of congenial effects, causal and constitutive: in the nature of the case, congenial, disposition-independent effects.

Thus your acts or deeds may give me pleasure or satisfaction, thereby scoring high in utilitarian terms. They may provide me with money or opportunity or other such resources. They may testify to my standing, increasing the esteem in which I am held in our society. And they may also do me good on the basis of constitutive linkages rather than causal connections. Your communication may constitute truth-telling, for example, thereby increasing my knowledge and capacity. And your action may constitute promise-keeping, thereby giving me the satisfaction of an associated claim.

But apart from these straightforward ways in which you may do me good, there are many ways of doing me good that involve disposition-dependent effects. You do me good in these ways, not just because of the acts or deeds you perform, but because of your actions or doings. You do me good because of the dispositions out of which you act: because of what you control for in producing your acts or deeds.

Respect

The example of respect serves to illustrate the possibility. You do me important good by almost any account of value in showing me respect. And, as we have seen, you show me respect only insofar as you exercise restraint in your dealings with me out of a suitable disposition. You do not act out of a motive of opportunism or whim or whatever, but out of a respectful disposition. You act out of a disposition to exercise restraint across a range of possible circumstances—you control for displaying restraint in those scenarios—including the circumstances that actually materialized.

Suppose I have a choice, within the sphere of the basic liberties, between doing either of two things: speaking out against the government, speaking in favor of the government, or remaining silent. And suppose that I choose to speak in favor of the government and that you

exercise self-restraint, allowing me to say my piece. Do you show me respect? Well, you do not show me respect if your restraint comes only of the fact that you are a government supporter yourself. You show me respect in your exercise of restraint only if it is the case that you would also have exercised restraint if I had chosen one of the other options: only if you controlled for the exercise of such restraint across those possibilities as well as in the possibility actually realized.

Putting this from my perspective rather than yours, I enjoy your actual respect only if you act in a certain way as things actually are—you exercise restraint—and only if you act a certain way—again you exercise restraint—under certain unactualized possibilities. The good of respect that I can enjoy at your hands requires you to exercise restraint, not just actually, but with a certain robustness: that is, across a certain range of possible variations on actual circumstances. But how can you provide me with this robustly demanding good: this good that requires things to be thus and so in non-actual as well as in actual circumstances? The obvious answer is: only by performing the required act—the exercise of restraint—in manifestation of a suitable disposition; only by adopting a respectful course of action.

Controlling for the exercise of restraint in the sense associated with respect means controlling for its realization even in scenarios that are quite unlikely to materialize. It means controlling for restraint in the scenario where I wish to speak against the government, for example, even if it is highly improbable that I should want to do so. Thus controlling for restraint is very different from maximizing expected self-restraint. If you wanted to maximize expected self-restraint in dealing with me, you could ignore scenarios that were almost certain not to materialize, such as the scenario where I wish to speak against the government. You would need to ready yourself to provide restraint across various scenarios only in the degree to which they were likely to materialize; to devote resources to unlikely scenarios would be utterly wasteful. Making restraint suitably robust is more demanding than making it suitably probable, then; controlling for restraint is more demanding than maximizing its expectation.

I now want to suggest that as the delivery of respect requires you to control for restraint in this sense, acting out of a suitable disposition, so the delivery of other important goods requires you to control for acts associated with them in the way that the act of self-restraint is associated

The Asymmetry of Good and Evil

23

with respect. There are any number of goods you can give me that, like respect, are robustly demanding in character: they require that you perform a certain sort of act, not just actually, but in some non-actual circumstances. Like respect, these are goods that you can put my way only by virtue of the dispositions out of which you act and so only by virtue of what you control for. The feature of these goods that is of particular importance, as we shall see, is that they require you to perform certain beneficial acts, not just when it is personally suitable, but also when this requires a certain frustration or even sacrifice on your part.

There are two categories where such goods figure prominently. One involves attachments like those of love and friendship, where you are bound to another person in their individuality, albeit under the aspect of a lover or friend or whatever. The second involves virtues of a substantive kind—for example, honesty or justice, tolerance or fidelity—that are targeted at another's need for information or acceptance or whatever. They are different from respect in the sense that the need involved is not conditional on the agent's making one or another choice: the need, should I choose X, not to be interfered with in that choice; the need, should I choose Y, not to be interfered with in that choice; and so on.

Love and Friendship

Take the good of love or friendship that is celebrated under almost any conception of human welfare. When you show me love or friendship, as we generally put it, you must perform a certain act or service: you must give me a form of care in the case of love, let us say, a form of favor in the case of friendship. But providing that actual service—performing a caring or favoring deed—is not enough for you to give me the good of love or friendship. After all, you might have performed it out of an opportunistic sense of your own self-interest or even out of a disinterested sense of what you are impersonally required to do. In order to provide me with love or friendship you must provide me with that service out of a disposition to do so across a range of possibilities, including possibilities where doing so impacts on your self-interest or frustrates certain cherished projects. You must provide me with the care or favor more or less robustly, not just contingently on its fitting well with your prudential or moralistic projects. It is by controlling for giving me care and favor in those possibilities, as well as in the actual scenario, that you give me love or friendship.

Thus consider the situation charted in Oscar Wilde's comedy, 'The Importance of Being Ernest'. Jack Worthing uses the pseudonym 'Ernest' on his visits to London, as he wishes to retain a certain anonymity in the big city. Under that pseudonym he attracts the attentions of Gwendolen, the cousin of his friend, Algernon, and she declares her love for him in response to his own confession of attachment. But does she really offer him love? In explaining her attachment, Gwendolen suggests that it is as contingent as his pseudonym: 'my ideal has always been to love some one of the name of Ernest', she says. 'There is something in that name that inspires absolute confidence'. And as if that were not sufficiently bewildering, she adds: 'The moment Algernon first mentioned to me that he had a friend called Ernest, I knew I was destined to love you'. Jack remonstrates with her, of course, explaining that he would much rather be called 'Jack'. But Gwendolen will have none of it, expounding with enthusiasm on the charms of 'Ernest'. 'It suits you perfectly. It is a divine name. It has a music of its own. It produces vibrations.'

The comedy of this situation derives from the fact that if Gwendolen is to count as giving Jack her love, then the care she purports to give him must not be contingent on something as incidental as his having a name that appeals to her. To give someone love is to give them care—to perform deeds of care—with a degree of robustness that rules out such sensitivity to personal taste. 'Love is not love', as Shakespeare puts it, 'Which alters when it alteration finds'. To love someone is to provide them with care across a range of possibilities that includes the possibility where their name changes, as well as various possibilities where there are other uncongenial changes in the beloved, or the care is more inconvenient for the lover in other regards. It is to provide them with care out of a disposition that ensures a high degree of robustness; in particular, a degree of robustness under which the care is not provided contingently on its fitting the precise tastes or advancing the personal interests of the lover.

We need not rule on the exact range of possibilities over which you must give me care in order to count as loving me, as we need not rule on the exact range of possibilities over which you must give me favor in order to count as my friend. Our common conceptions of love and friendship are vague on those margins. You will love me to the extent that you give me suitably robust care and you will be treat me as a friend

The Asymmetry of Good and Evil

25

to the extent that you give me suitably robust favor. You will love me or treat me as a friend to the extent that you control in a relatively benevolent measure for giving me care or favor across those various possible situations where the considerations that trigger love or friendship obtain. But however vague at the margins, the possible situations across which your care or favor must persist include many where there is a certain, self-sacrificing cost to you in providing it. Love and friendship require a disposition to care for the lover and favor the friend, even when doing so is inconvenient or goes in other ways against the grain.

By all accounts you do me good insofar as you show me love or friendship, as by all accounts you do me good insofar as you show me respect. And that means that by common consensus you do me good insofar as you act out of the dispositions associated with love and friendship, controlling in a benevolent measure for giving me the associated care or favor. The love or friendship you provide comes about as a disposition-dependent effect of your providing care and favor. It is an effect of your action but not, in our sense, of the act that you produce in the course of taking that action.

Honesty, Justice, and other Virtues

This claim in the case of attachments like love and friendship is parallel to a similar claim that can be made in the case of various virtues, in particular virtues that govern how people treat one another. As you do me good in giving me love or friendship so, by our received ideas, you do me good in giving me the benefit of your honesty or justice, your tolerance or acceptance, or your fidelity as a promise-keeper. And as love or friendship require you to control in a suitable measure for certain good effects like care and favor—to produce them, in particular, with a suitable degree of self-sacrificing robustness—so these virtues require you to control in a corresponding degree for parallel effects.

The point is readily illustrated. In order to give me honesty you must tell me the truth, in order to give me justice you must satisfy my claims against you, in order to give me tolerance or acceptance you must treat me decently, and in order to give me fidelity you must keep your promises. But it will not be sufficient in any of these cases that you contingently provide such truth-telling or claim-satisfaction, decent treatment or promise-keeping. You must perform the required acts or deeds, not just as things actually happen to be, but also across a range of

unactualized possibilities, including possibilities where it is relatively costly for you to do so. You must control in a relatively benevolent degree for realizing the good options in question.

As in the case of attachments like love and friendship, our common conceptions of the virtues will dictate an indeterminate but demanding threshold that you must meet in controlling for the appropriate effects if you are to count as honest or just, tolerant or accepting or faithful. And as in the other case, we need not worry particularly about where this threshold lies. The important point to note is that the threshold is sufficiently demanding to imply that you do not count as honest unless you are prepared to tell the truth when it is inconvenient, you do not count as faithful unless you are prepared to keep a promise that you would prefer not to have made, and so on. This means that in the area of virtue, as in the areas of respect and attachment, we routinely think that doing good requires that you control for taking good options, and do not just happen to produce them or produce them for contingent reasons of convenience or interest. You do me good as a result of the disposition-dependent effects of the actions you perform.

3. DOING EVIL

Do Evildoers Exercise Malevolent, Self-sacrificing Control?

We have seen that while doing good may sometimes mean just realizing good options—doing good deeds or acts, as we may say—it routinely means controlling for their realization. It means ensuring for a range of possibilities, not just for the actual situation on hand, that you take those options, where the possibilities include some in which the performance runs against your self-interest. This observation raises a straightforward question about doing evil as distinct from doing good—specifically, for doing evil as distinct from doing good to another human being. Does doing evil require controlling for evil deeds in the way in which doing good often means controlling for good deeds? In particular, does it require controlling for evil deeds even across possibilities where those deeds impose a self-sacrificial cost on the agent?

There can be little doubt but that the effects in virtue of which we might say that you do evil, imposing some costs on me, can include disposition-dependent effects. If you hurt me out of thoughtlessness that is not so bad as hurting me out of indifference, which in turn is not so

The Asymmetry of Good and Evil

27

bad as hurting me out of enmity. But the question in the case of doing evil is whether there are cases parallel to those illustrated by attachments and virtues where the evils involved require controlling in a similar, self-sacrificing measure—controlling malevolently, not benevolently—for the performance of harmful deeds. Do various forms of evil-doing require that you should perform certain evil deeds with a similar degree of robustness? Do they require in the same way, not only that you actually perform those evil acts, but that you would perform them across a range of possibilities, including possibilities where the performance imposes a cost in self-interest?

Doing good deeds with the robustness required by attachment, virtue or respect has a somewhat heroic character, since the control involved demands some self-sacrifice on your part; it requires that you set aside your self-interest in a certain measure, being prepared to do those deeds even when that impacts on your own interests. The question here is whether there are robustly demanding evils, parallel to the robustly demanding goods of attachment, virtue and respect, that require a similar sort of self-sacrifice and a correspondingly heroic character.

Answer: Not in General, Only in Exceptional Cases

The answer, intuitively, is that there are not many evils of this robustly demanding kind. When you do evil you certainly perform an evil deed, realizing an option that has negative effects on me, whether of a disposition-independent or disposition-dependent character. But the sort of evil pursued does not often require that you should control in a heroic degree for harming me in that way. It does not often require that you should be willing to impose the harm, even in scenarios where it is costly for you to do so.

There are certainly a limited number of cases where evil-doing involves controlling malevolently for imposing a harm on another, whatever form that harm may take. One example might be the case of acting out of envy or spite where you are required by the nature of the evil you pursue to hurt another, even if it should cause you great loss as well. Another might be the case of revenge where that goal requires you without heed for your own well-being—say, without heed for the retaliation you may trigger—to do harm to another. And a third might be the sort of racial hatred promoted in Himmler's 1943 Potsdam speech, when he encouraged his troops to maintain their genocide against the Jewish

people at whatever cost to their own feelings or sensitivities. But the striking thing about these cases is that they are so extreme. In each example, you are required as an evil-doer to be ready, in an old phrase, to cut off your nose to spite your face. You are required, as that phrase suggests, to behave in what is taken to be a self-defeating, even irrational manner.

To be an heroic evil-doer of the kind envisaged here would be to espouse the sort of attitude ascribed to Aaron in the final scene of Shakespeare's *Titus Andronicus*: 'If one good deed in all my life I did, I do repent it from my very soul'. This kind of maleficence is demonic in character, fitting the slogan of Milton's Satan: 'Evil, be thou my good'.⁷ But the heroic character of the exercise suggests that it is likely to be very rare. There are lots of bad things that we human beings do to one another, but there are few plausible evils that require self-sacrificial action (Baumeister 1997; Russell 2006). Doing evil is generally a much more banal phenomenon—even among Nazis, as Hannah Arendt (1963) famously suggested.

Hobbes (1994 [1668], Ch 15) recognizes the non-heroic aspect of evil-doing, in particular injustice, when he says in *Leviathan* that whereas a just man is 'he that taketh all the care he can that his actions may all be just, an unjust man is he that neglecteth it'.⁸ The unjust man's will is framed, as he goes on to explain, not by desire for injustice as such, but 'by the apparent benefit of what he is to do'. On this account, the evil of injustice does not consist in imposing damages on others out of a pure, selfless desire to do so. On the contrary, it consists in the failure to honor people's claims of justice in how you treat them—the failure that consists in an intentional or at least negligent action—because of a self-seeking search for reward or benefit.⁹

4. EXPLAINING THE ASYMMETRY BETWEEN GOOD AND DOING EVIL

The Role of Standards

What Hobbes directs our attention to is the fact that whereas doing good often means controlling for the performance of good deeds, even at some

⁷ The line in this paragraph gives us reason to be reluctant to endorse Thomas Nagel's (1980, 132) claim: 'To aim at evil, even as a means, is to have one's action guided by evil'.

⁸ Charles Pigden (1988, 29) puts the point nicely in endorsing a view he finds in Aquinas: 'the just man is not one who makes a policy of injustice — merely one who does not make a policy of justice'.

⁹ For similar lines of thought in the medieval period, see (Anselm 2000). I am grateful to Calvin Normore for pointing this out.

The Asymmetry of Good and Evil

29

cost to yourself—*taking all the care you can* that your deeds be good—doing evil has a different character. It is more likely to consist, not in controlling with a corresponding robustness for the performance of evil deeds, but in performing evil deeds as a means to satisfying your self-interest or as a side-effect of satisfying your self-interest. The person who does good often controls for performing good deeds, even at some cost to their self-interest. The person who does evil typically controls for the satisfaction of their self-interest, where the cost of doing so involves imposing some harm on another: doing an evil deed.

There should be no surprise about this asymmetry between doing good and doing evil. The standards that we hold up for one another, including standards of respect, attachment and virtue, direct us to robustly demanding goods, in particular goods that often demand a degree of self-sacrifice. And there are no standards of the same kind that would direct us to evils of an equally demanding character: evils such that doing them would require you to control malevolently, even self-sacrificially, for imposing certain costs on others.

Why do we have standards of goodness that we effectively promulgate, being ready to approve of conformity to the standards and to disapprove of deviance—and being ready to do this on public grounds, not just on grounds of private interest? We are creatures of limited altruism who have to live in a world of scarce resources and these circumstances often put us in a situation where we each do better in self-interested terms to the extent that we all abide by standards that limit our direct pursuit of self-interest. We do better by sticking to the standards prescribed for lovers and friends, since we can thereby build mutually advantageous relationships rather than finding ourselves isolated from one another. We do better by sticking to standards of honesty and justice and fidelity in our dealings with one another, even with strangers, because this makes for a peaceful world in which we can plan our lives with a degree of assurance.

There are many stories about how we human beings come to identify such standards and implement them more or less effectively as social norms but we need not look at the relative merits of those accounts here. It suffices for our purposes to assume that, however their emergence and maintenance is explained, the standards are commonly recognized, with cultural variations, across most societies. In each of our communities, and even across different communities, we acknowledge collectively

beneficial standards of good behavior to which we hold one another, approving of conformity and disapproving of non-conformity. And we hold one another to these standards in full recognition that they are often individually burdensome, requiring people to put the good of acting as the standards require above their immediate self-interest.

It is no surprise that given the presence of suitable standards, we expect people who count as doing good to be disposed to comply with those standards and to act out of that disposition. We expect them to control for acting as lovers or friends, or as honest, just and faithful brokers, when such action is required by the needs of others, according to the shared standards. And of course we expect them to do this even across possible scenarios where their immediate self-interest is frustrated; after all, the standards are designed precisely to overcome the problem of limited altruism.

The existence of standards that are designed to cope with our limited altruism makes sense, then, of why we associate various forms of doing good with controlling benevolently for good acts like those of care and favor, truth-telling and claim-satisfaction, promise-keeping and self-restraint. But the striking thing on the side of evil is that there are generally no parallel standards there. At least within the organization of a given society, there is no basis for wanting people to overcome their immediate self-interest in imposing certain costs on others. It is not as if such a pattern would be collectively beneficial. And so there are no standards of evil-doing that serve the role associated on the side of good with standards of attachment, virtue and respect. If there are any exceptions, they involve standards governing people's behavior towards outsiders or those cast as outsiders. Himmler's promulgation of racial hatred offers a chastening example.

If most societies promote standards of good but few if any societies promote standards of evil, it is understandable that while we have relatively heroic standards for doing good—those that require controlling benevolently, even self-sacrificially, for performing certain good deeds—we do not have heroic standards for doing evil: standards that would require controlling malevolently, even self-sacrificially, for performing certain evil deeds. What evil-doing is likely to be taken to involve, in the presence only of standards for good, is the imposition of costs in a way that breaches those standards: an imposition of costs, presumably, that is prompted by self-interest. And that is precisely what

The Asymmetry of Good and Evil

31

Hobbes suggests when he says that to do justice is to take all possible care that you act justly whereas to do injustice is to neglect the requirements of justice—to act unjustly—because of the benefit that that promises for yourself. To do justice, and to do good of any of a variety of kinds, is to control benevolently, even self-sacrificially for doing just or good acts. To do injustice, and to do evil of any of a range of forms, is to breach the requirements of good because of controlling for self-interest.

Historical Antecedents

The line emerging from these considerations does not find support only in the thought of Hobbes. It fits with the long tradition of thinking, associated with Augustine and Aquinas, that rejects the Manichaean idea that good and evil are of equal metaphysical importance and that the moral life involves a struggle between the two principles. Without delving into the metaphysics associated with the position, it should be clear that the observations we have rehearsed may give some plausibility to this line of thought.

The primary struggle in moral life is to live up to the demands of our standards of good, such as the standards associated with various attachments and virtues. What it means for good to lose out, and evil to triumph, is merely that you fail to do good: you act in a way that does not live up to the standards of good, being tempted away from the high path charted in those standards by the temptations of immediate self-interest. Thus good is the positive goal and evil means acting in a way that is contrary to that goal: failing in that sense to live up to the goal. In the received Latin phrase, *malum est privatio boni*; evil is merely the absence of good, not something positive in itself.

Not only do our observations offer some support for this anti-Manichaean way of thinking. They may also help to make sense of another medieval principle of moral thought, which received particular prominence in the work of Aquinas. This principle holds that anything desired is desired under the aspect of good: *quidquid appetitur sub specie boni appetitur*. It is often presented as a doctrine governing the very possibility of intelligible desire and intentional action and it is often criticised in that role (Velleman 1999). But it may be better taken as an affirmation that whereas we often desire the good of others for its own sake, controlling benevolently for conferring associated benefits, we generally seek the bad of others in order to promote our own self-

interest or at least on the proviso that our own self-interest is preserved. We do not desire the bad of others for its own sake but only as a means or a complement to our own good.

5. THE KNOBE EFFECT

Joshua Knobe (2003; 2006; 2010) introduced the effect now named after him by testing a simple vignette on a variety of subjects, identifying a difference in their willingness to describe a would-be good action and a would-be bad action as intentional (see too Pettit and Knobe 2009). Knobe's vignette involves two stories about the chairman of a company board. First story: asked to consider a program that he is told will increase profits but harm the environment, the chairman answers that the company's only concern is to increase profits—the environment does not matter—and that the program should be put in place. Second story: asked to consider a program that he is told will increase profits and help the environment, he answers again that the company's only concern is to increase profits—the environment does not matter—and that the program should be implemented.

Does the chairman intentionally harm the environment when he acts as in the first story? Does he intentionally help the environment when he acts as in the second? Most people, so it turns out, think that he intentionally harms the environment in the first story but that he does not intentionally help the environment in the second. The Knobe effect consists in the fact that this asymmetry in the ascription of intentionality characterizes the responses of most human beings in this sort of vignette.

The effect is challenging, because it suggests that the folk psychology on which we rely in distinguishing between cases where people do things intentionally and cases where they don't is warped by our evaluative viewpoints. The idea would be that despite the fact that there is nothing of psychological significance to distinguish the position of the chairman in the two scenarios, still we find a psychological difference there. We draw a psychological distinction, it appears, not in response to considerations that are germane to a presumptively explanatory enterprise, but rather in response to irrelevant, ethical considerations. We practice a psychology that is essentially value-dependent.

Knobe himself appears to be willing to live with this value-dependence, for he explains the effect by hypothesizing that ascriptions

The Asymmetry of Good and Evil

33

of intentionality are sensitive to moral considerations, and cannot be taken to be exclusively or even primarily explanatory. But it turns out that our observations about the asymmetry of good and evil make sense of people's intuitions about ascriptions of intentionality, without resort to any such hypothesis.¹⁰

Why do we readily think in the first story that the chairman does intentionally harm the environment but in the second story that he does not intentionally help the environment? The asymmetry that we have noticed between doing good and doing evil supports a natural suggestion: that helping involves controlling for compliance with environmental standards, whereas harming does not; that this makes it harder to count as helping rather than harming; and that consequently there are stricter requirements that the chairman must meet if he is to be cast as intentionally helping the environment than if he is to be cast as intentionally harming it.

Most of us think that there are standards that formulate the preconditions for a healthy environment. Indeed the very questions raised in the vignette presuppose an assumption of this kind; without that assumption the idea of helping or harming the environment would otherwise make little sense. What is to count as helping the environment, against the background of such standards, and what is to count as harming it? It will certainly be enough to count as harming the environment that you breach environmental standards, say for reasons of self-interest; you do not have to be an environmental vandal who controls self-sacrificially for flouting those standards. But it will hardly be enough to count as helping the environment that you happen to conform to the standards, again for reasons of self-interest. Here the natural suggestion is that to help the environment—to do good in that respect—it is essential that you control for conformity to environmental standards, even at some loss to yourself (see too O'Brien forthcoming).

This means that if the chairman is to count as intentionally helping the environment he has to cross a high bar, intentionally controlling for

¹⁰ The line taken here parallels in some measure the approach championed by Richard Holton (2010) although his approach is defended in quite different terms. For an alternative approach, but with some affinities, see (Hindriks 2013). Quite independently, Lillian O'Brien (forthcoming) has found empirical support for an idea close to that defended here: viz., that respondents are triggered in Knobe-like experiments by the different conceptions that they hold of the requirements of helping, on the one hand, and harming on the other.

conformity to environmental standards. And obviously the chairman in the first vignette does not reach that bar, so that people are quite right to say that he does not help the environment intentionally. If the chairman is to count as intentionally harming the environment, however, he merely has to cross the bar of breaching those standards intentionally. And that is easily met: breaching those standards does not require controlling for breaching them, in the spirit of a vandal; and, in ordinary usage, it is enough for him to breach those standards intentionally that he is aware of what he is doing (Scanlon 2008; Pettit 2015).¹¹ Thus people are quite right in this case to assert that the chairman does harm the environment intentionally. There is no mystery about the difference of attitude they assume in the two cases.

Knobe and his colleagues tried many variations of the original vignette on people, with broadly similar results. Instead of asking whether the chairman intentionally helped or harmed the environment, for example, people were asked whether he decided to help or decided to harm, whether he favored helping or favored harming, whether he advocated helping or advocated harming.¹² Unsurprisingly from our point of view, the results retained much the same asymmetry as in the original case. The chairman certainly doesn't decide to help, or favor or advocate helping, given that helping requires controlling for conformity to environmental standards. But it is quite plausible to think that he decides to harm, or even favors or

¹¹ For the record, I think that the requirements for doing X intentionally are weaker than the requirements for acting with the intention to X and that they in turn are weaker than the requirements for forming the advance intention to X. In order to form the advance intention to X it is necessary to do more than just act out of a desire to X, taking whatever steps are believed to be suitable; you have to let X-ing at the appropriate time assume the place of a relative fixture in your decision-making (Bratman 1987). In order to act with the intention of X-ing you do not have to act on the basis of a pre-formed intention; you need only act out of a desire to X, adopting the steps you believe to be appropriate. And in order to do X intentionally you do not have to act out of a desire to X; you need only act out of a desire to Y in a case where you foresee that Y-ing will involve X-ing and you are prepared to live with that result: as it may be, with that cost.

¹² They also asked whether the chairman can be said to desire harming or to intend to harm—that is, to act with the intention of harming. To say either would be to suggest that he is attracted to the option under all its aspects—see the previous footnote—whereas to say he harmed intentionally is consistent with his desiring or intending the action under another aspect only—say, as a way of making profits; in this case he merely foresees and accepts that the harm will occur (Scanlon 2008; Pettit 2015). It is not so surprising, then, that in these cases the results are less decidedly asymmetrical: that people tend to say that he did not desire or intend to harm, as he did not desire or intend to help (Pettit and Knobe 2009).

The Asymmetry of Good and Evil

35

advocates harming, given that harming requires only breaching environmental standards.

The challenge raised by the Knobe effect is to vindicate the explanatory role of ascribing intentionality by showing that the ascription is not value-dependent. Our story meets that challenge, because it traces the effect to the different sorts of actions represented by helping on the one side and harming on the other. We do not take different lines in the two cases because one action has positive value, the other negative. We take different lines because of a difference between the requirements that must be satisfied if an action is to count as helping or as harming.

But is the difference we trace still value-dependent insofar as helping is a positive action, harming a negative one? No, it is not. The fact that there is a high bar for helping and a low bar for harming is what explains the difference in our ascriptions of intentionality and it is incidental that the explanation for the difference in the bar required goes back to matters of good and evil. Thus we would expect to find an analogue of the Knobe effect wherever there is a high-low difference in the bar to be crossed, and not just when that difference has moral significance.

Take the difference in the bar associated with following the rules of grammar and breaking the rules of grammar. And now imagine a vignette similar to Knobe's. The chairman is told in one story that the advertising branch has come up with a great company slogan but one that is ungrammatical and in the other that it has come up with a great slogan and one, moreover, that is grammatical. In each case he says: I don't mind about the grammar; if it's a great slogan, let's use it. Does he intentionally breach the rules of grammar in the first story? Surely he does. Does he intentionally follow the rules of grammar in the second story? Surely he doesn't. What then is the difference? Clearly, I would say, the fact that there is a high bar for following the rules of grammar—you must control for conforming to them—and a low bar for breaching them: you must certainly breach the rules but you need not control for breaching them.

We saw in the first section that the things we do have both causal and constitutive effects and that among their constitutive effects are ones that depend on the dispositions out of which we act. We saw in the second section that these disposition-dependent effects are very important in doing good to others, as when we act out of attachment, virtue or respect; in particular, when we do so in neglect of our own self-interest.

We saw in the third section that in doing ill to others, such disposition-dependent effects are not of the same importance and, in particular, that most forms of evil-doing do not require us to neglect our own self-interest in the same way. But there is no mystery in this asymmetry, as we saw in the fourth section. The explanation is that doing good generally involves living up to certain standards that guard against limited altruism whereas doing evil generally involves acting in breach of such standards, not living up to any counterparts on the dark side.

With these observations in place, we can now see that the Knobe effect, which is so prominent in recent discussions of moral psychology, is not mysterious either. If helping the environment means acting up to environmental standards, and harming the environment means failing to act up to those standards, then it is much harder to count as helping than it is to count as harming. And so it is unsurprising that the company chairman in Knobe's vignette can be thought to harm the environment intentionally but not to help the environment intentionally. This asymmetry simply reflects the more general asymmetry between helping and harming, doing good and doing evil.

Acknowledgments

I am grateful for helpful discussions of the material in this paper when I presented it as an Inaugural Lecture at the Australian National University in May 2013, as the Dan and Gwen Taylor lecture at the University of Otago in March 2014, and as the Kendrick lecture at the Arizona Workshop in Normative Ethics in January 2014. I am also grateful for very useful comments received from an anonymous referee, and from the editor, Mark Timmons.

REFERENCES

- Anscombe, G. E. M. (1957). *Intention*. Oxford, Blackwell.
- Anselm of Canterbury. (2000). 'The Fall of the Devil', in *The Complete Philosophical and Theological Treatises of Anselm of Canterbury*. J. Hopkins and H. Richardson (eds). Minneapolis, Arthur J Banning Press.
- Arendt, H. (1963). *Eichmann in Jerusalem: A Report on the Banality of Evil*. New York, Viking Press.
- Baumeister, R. (1997). *Evil: Inside Human Violence and Cruelty*. New York, Henry Holt.
- Bennett, J. (1995). *The Act Itself*. Oxford, Oxford University Press.
- Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Cambridge, MA, Harvard University Press.

The Asymmetry of Good and Evil

37

- Davidson, D. (1980). *Essays on Actions and Events*. Oxford, Oxford University Press.
- Goldman, A. (1970). *A Theory of Human Action*. Englewood Cliffs, NJ, Prentice-Hall.
- Herman, B. (2011). 'A Mismatch of Methods', in *Derek Parfit: On What Matters*, vol. 2. S. Scheffler (ed.). Oxford, Oxford University Press: 83–115.
- Hindriks, F. (2013). 'How to Explain the Knobe Effect and its Relatives'. *Mind and Language* 28.
- Hobbes, T. (1994 [1668]). *Leviathan*. E. Curley (ed.). Indianapolis, Hackett.
- Holton, R. (2010). 'Norms and the Knobe Effect'. *Analysis* 70: 1–8.
- Knobe, J. (2003). 'Intention, Intentional Action and Moral Considerations'. *Analysis* 64: 181–7.
- Knobe, J. (2006). 'The Concept of Intentional Action'. *Philosophical Studies* 130: 203–31.
- Knobe, J. (2010). 'Person as scientist, person as moralist'. *Brain and Behavioral Sciences* 33: 315–65.
- Korsgaard, C. (2009). *Self-Constitution: Agency, Identity, and Integrity*. Oxford, Oxford University Press.
- Nagel, T. (1980). 'The Limits of Objectivity', in *The Tanner Lectures on Human Values*, vol. 1. S. McMurrin (ed.). Salt Lake City, University of Utah Press.
- O'Brien, L. (forthcoming). 'Side-Effects and Asymmetry in Act-Type Attribution'. *Philosophical Psychology*.
- Pettit, D. and J. Knobe (2009). 'The Pervasive Impact of Moral Judgment'. *Mind and Language* 24: 586–604.
- Pettit, P. (1995). 'The Virtual Reality of Homo Economicus'. *Monist* 78: 308–29. Expanded version in U. Maki (ed.), *The World of Economics*, Cambridge University Press, 2000; reprinted in P. Pettit (2002), *Rules, Reasons, and Norms*, Oxford, Oxford University Press.
- Pettit, P. (2015). *The Robust Demands of the Good: Ethics with Attachment, Virtue and Respect*. Oxford, Oxford University Press.
- Pigden, C. (1988). 'Anscombe on "Ought"'. *Philosophical Quarterly* 38: 20–41.
- Ross, D. (2002 New Edition). *The Right and the Good*. Oxford, Oxford University Press.
- Russell, L. (2006). 'Dispositional Accounts of Evil Personhood'. *Philosophical Studies* 149: 231–50.
- Scanlon, T. M. (2008). *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, MA, Harvard University Press.
- Velleman, D. (1999). 'The Guise of the Good'. *Nous* 26: 3–26.