

The paradox of deontology, revisited

ULRIKE HEUER

It appears to be a feature of our ordinary understanding of morality that we ought not to act in certain ways at all. We ought not to kill, torture, deceive, break our promises (say)—exceptional circumstances apart.¹ Many moral duties are thought of in this way. Killing another person would be wrong even if it achieved a great good, and even if it led to preventing the deaths of several others. This feature of moral thinking is at the core of deontological ethics. But while it is also part and parcel of ordinary moral thought, it seems puzzling, because it requires an agent to act (or to abstain from acting) in certain ways, even if by violating the duty she could achieve a better outcome in terms of the very restriction that she is required to heed.

If the explanation of the duty not to kill were that it preserves human life it would be hard to understand why killing is not morally permitted when it leads to preserving a greater number of lives. Moreover, why should we not violate a restriction if doing so leads to fewer violations of that very restriction? There are two questions here. (1) Why not violate a restriction whose point it is to preserve lives, if doing so leads to preserving a greater number of lives? (2) Why not violate a restriction if doing so minimizes the violation of that very restriction? The latter question concerns what is sometimes called ‘the paradox of deontology’ or the problem of ‘minimizing violations’. The former may concern the so-called trolley cases, for instance. After all, a trolley hurtling towards a number of innocent people is not about to violate the moral restriction

¹ Many deontologists are ‘threshold deontologists’, meaning either that consequences of a certain kind—catastrophic ones—must be avoided even at the price of violating restrictions, or that the restriction can be defeated by consequences whose disvalue is out of proportion with the particular value that the restrictions protect.

on killing innocent people, as it is not the kind of thing that can be subject to such a restriction. In this paper I will be concerned only with (2): the paradox of deontology.

To put it in Nozick's terms: 'How can a concern for the non-violation of C [i.e. some deontological constraint] lead to refusal to violate C even when this would prevent other more extensive violations of C?'²

Samuel Scheffler, puzzled by the same idea, put it in terms of 'agent-centred restrictions', because the restriction requires an agent (not) to act in certain ways, contrasting thereby with a requirement that certain actions not be done. The restriction is thus an agent-relative rather than an agent-neutral one.³ Scheffler ended up rejecting the idea of agent-centred restrictions, as even on close investigation the puzzle would not dissolve. Here is his description of the structure of agent-centred restrictions (= ACR):

Suppose that if agent A_1 fails to violate a restriction R by harming some undeserving person P_1 , then five other agents $A_2 \dots A_6$, will each violate restriction R by identically harming five other persons, $P_2 \dots P_6$, who are just as undeserving as P_1 , and whom it would be just as undesirable from an impersonal standpoint to have harmed.⁴

If there are agent-centred restrictions it can be wrong for A_1 to violate R, even in this situation. Scheffler maintains that ACRs have 'an air of paradox' to them, as they violate what he calls maximizing rationality.⁵ ACRs appear to clash with the idea of maximizing rationality, because they require an agent to choose the option that worse achieves her goal of not harming people.

Let me call this, following common parlance, the paradox of deontology, even though it is not strictly speaking a paradox. Exploring

² R. Nozick (1974), p. 30.

³ I will not discuss the distinction between agent-relative and agent-neutral requirements or reasons in any detail, but I will return to it briefly in Section III, and explain why the agent-relativity of the relevant reasons would not be sufficient for establishing agent-centred restrictions. Just as a general explanation, an agent-relative restriction on killing requires of any agent that she must not kill; an agent-neutral one requires of any agent to see to it that there be no killings. A note on terminology: I will use the expressions 'agent-centred restrictions' and 'deontological restrictions' interchangeably.

⁴ S. Scheffler (1982), p. 84.

⁵ Scheffler (1985), quoted from *The Rejection of Consequentialism*, second edn., p. 143.

whether there is really a problem here, and how it might be solved, is the main objective of this paper.

I will begin by explaining the puzzlement about agent-centred restrictions in more detail. I will then discuss promising as an example of a kind of action which is thought to be subject to agent-centred restrictions. It will turn out that they are surprisingly easy to explain. The difficult case is killing—and that should put us on guard. I am going to argue that the problem is not with the very idea of deontological restrictions, but with puzzling aspects of the ethics of killing.

I. THE PARADOX OF DEONTOLOGY

Some philosophers find deontological restrictions not at all puzzling. If you are under the spell of consequentialist thinking, or in the grip of ‘maximizing rationality’, there is a puzzle, they maintain. But you only have to shed those influences and the puzzle will vanish with them. According to ‘maximizing rationality’, if one of two available options achieves an agent’s goal better than the other, a rational agent ought to choose the one that is better. As Scheffler put it:

The core of this conception of rationality is the idea that if one accepts the desirability of a certain goal being achieved, and if one has a choice between two options, one of which is certain to accomplish the goal better than the other, then it is, *ceteris paribus*, rational to choose the former over the latter.⁶

Deontological restrictions violate this requirement of maximizing rationality, because, as far as the goal of not killing or harming the innocent is concerned, five innocent people killed (or harmed) compared to one killed (or harmed) is clearly the worse option, and therefore the one that the agent ought to prevent. This presupposes, however, that it is the impartial value of realizing some goal which serves as the rationale for the agent-centred restriction. But the deontologist ought not to accept this. As Scanlon points out,⁷ the ‘maximizing rationality’ claim may be

⁶ Scheffler (1985), quoted from *The Rejection of Consequentialism*, second edn., p. 143.

⁷ Scanlon (1998), discussing Scheffler’s view (pp. 82ff) writes: ‘the argument . . . seems to rely on the assumption that the negative intrinsic value attached to morally undesirable actions is impartial—that is to say, that an action that has this disvalue is something that everyone . . . has reason to prevent’ (p. 83). And further: ‘[I]f one has adopted a goal (and does not have reason to reconsider its adoption), then one has, other things being equal, reason to

correct, but the deontologist will reject that the rationale of agent-centred restrictions is that they serve the goal of not killing the innocent. Thus there is no puzzle.

But the optimism is premature. Even if ACRs are not supposed to be explained by the goal of not killing the innocent, there has to be some other explanation.⁸ We could start by presenting ACRs in terms of the rights and duties that they involve—a language which is often taken to be more congenial to the deontologist's concern. Take the following description: 'A person (P_1 in ACR) has a right not to be killed, even if it would prevent the killing of five others. That P_1 has a right not to be killed implies that A_1 has a duty not to kill P_1 , no matter what the consequences or unless they are catastrophic.' The deontologist might claim that we can ground such a right in a more fundamental moral concern: the respect we owe to others, because of their value and dignity as persons.⁹ But how about the respect we owe the five? Of course they too are deserving of the same respect, and they too have a right not to be killed. But if A_1 were morally permitted to violate P_1 's right in order to help prevent A_2 's violation of the rights of $P_2 \dots P_6$, then we would have to interpret P_1 's right not to be killed differently. She does *not* have a right not to be harmed or killed, unless failing to do so would be catastrophic (call this the ACR-interpretation of rights). She only has a right not to be harmed or killed, unless failing to do so would prevent a greater number of similar rights violations (call this the MV-interpretation¹⁰ of rights).^{11,12} Why is the ACR-interpretation the right one?

prefer actions that better accomplish it. Taken as an observation about one common but nonetheless distinctive element in practical thinking, this claim is quite correct' (p. 85). But, Scanlon concludes, it is a misunderstanding to believe that the deontologist is one who has adopted the goal of preventing a certain kind of disvalue from occurring.

⁸ I take this to be Scheffler's point: ACRs cannot be explained by maximizing rationality—but what then would explain them? Would their explanation, whatever it turns out to be, be compatible with maximizing rationality?

⁹ Frances Kamm (2007), for instance, regards the 'moral status' of a person as an expression of her value: '... my account highlights an agent-neutral value: the high degree of inviolability of persons that lies at the base even of an agent-relative duty' (p.29). Similar in her 'Moral Status', in Kamm (2007), chap. 7.

¹⁰ 'MV' for 'minimizing violations'.

¹¹ There are, of course, other exceptions as well: a person may have no right not to be killed if the case is one of self-defence, or just war, for instance.

¹² Frances Kamm (2007) argues that this would be a lesser right. She claims that what is impermissible or impermissible to do to a person fixes her moral status. The moral status of rational persons is very high. If it were morally permissible to violate a person's right not to be

The MV-interpretation of rights is not committed to consequentialism: it does not require that an agent should always perform the action that produces the best outcome. MV is silent on outcomes in general. It only requires minimizing rights violations. The MV-interpretation is an alternative both to ACR and to consequentialism, and it is not obvious why it ought to be rejected in favour of the ACR-interpretation. This is another way of articulating what seems puzzling with ACRs: favouring the ACR-interpretation over the MV-interpretation is, without further argument, unjustified.

But let us just grant that (for whatever reason) the MV-interpretation of rights ought to be rejected, and accept the ACR-interpretation. (In Section IV I will present an argument why this is indeed correct.) Even so, there is yet another way of explaining why deontological restrictions seem puzzling. Even if we have an ACR-involving right not to be harmed or killed, is not there also a duty to help and protect people from coming to harm if we can? Some philosophers flatly deny this;¹³ others claim that such a duty is at any rate weaker than the duty not to violate the right not to be harmed.¹⁴

I will assume that there is a duty to help others and protect them from coming to harm—whether or not it is weak is a different matter.¹⁵ It

killed if it prevents a greater number of rights violations, her status would have to be lower than it actually is. This argument attempts to show that the ACR-interpretation of rights is correct and the MV-interpretation ought to be rejected. But I am doubtful about its success, because it seems to beg the question: A person's moral status is determined by the moral propositions that are true of her (for example, that she has an ACR-involving right not to be killed), and the moral proposition is true because it correctly reflects the moral status of a person. But we do not seem to have any independent purchase on the value of persons that would allow us to explain a person's status (and in turn why respecting it requires her having ACR-involving rights). We cannot determine a person's status independently of knowing what her rights are. Of course, Kamm may nonetheless be correctly articulating our (pre-theoretical) view of the status of persons, and their value. But there is no argument here why the moral status of rational beings requires that they have ACR-involving rights. The claim is just that this is how it is. As far as I can see, Kamm does not show that the ACR-interpretation is correct, but she presupposes that it is.

¹³ Perhaps some libertarians such as Narveson (2003). They may, however, like Nozick (1974), only wish to maintain that the state ought not to enforce this duty, rather than that there is none.

¹⁴ Philippa Foot (1967) writes: '[t]o refrain from inflicting injury ourselves is a stricter duty than to prevent other people from inflicting injury, which is not to say that the other is not a very strict duty indeed' (p. 29).

¹⁵ While this claim is not uncontested, it is nonetheless common ground between adherents of otherwise very different theories. Compare Peter Singer's claim (1972): '... if it is in our

seems obvious to me that if you can prevent a person from being killed or harmed at little or no cost to yourself, you must do so. If all it takes to save a life is to call an ambulance, for instance, failing to do so is a grave moral fault. But even if the costs are considerably higher, it still seems to me that you have a duty to help. Furthermore, those who defend ACRs often accept that we have such duties. Thus, the question is not whether a bystander has a duty to help the five who are about to be killed by someone else—I will assume that she does—but whether the reason to comply with this duty is defeated if it requires killing an innocent person.

A simple consideration supporting the claim that there is at least some duty to help if one can is the following: the ACR-involving right not to be harmed is supposed to be grounded in the value and dignity of persons. It is because persons are of value that they are protected by such rights. But how could the value and dignity of persons ground a right not to be harmed, but fail to ground a duty to help and protect?¹⁶ After all, we have such duties with regard to all other things of value. We have not only a duty not to destroy a valuable artwork, but also a duty to protect it when it is threatened to be destroyed. It would seem utterly bizarre to think that while you should not throw acid at Rembrandt's *Nightwatch*, if you see someone else setting out to do so (and you can prevent it) it is none of your concern. The same holds, *mutatis mutandis*, for animals, other artifacts, and perhaps for natural beauty. It seems that the duty not to harm and the duty to protect normally go hand in hand. Why would this be different when it comes to persons?

But if there is an obligation to help persons and protect them from coming to harm, then A_1 is not *only* under an obligation not to violate P_1 's right not to be killed, but she also has a duty to help $P_2 \dots P_6$, and it is by no means obvious how *this* conflict ought to be resolved. Thus, the very same value, the special value of persons, gives rise to potentially conflicting

power to prevent something very bad from happening, without thereby sacrificing anything else morally significant, we ought, morally, to do it', and an Aid Principle, put forward by T. M. Scanlon (1998): '... if you are presented with a situation in which you can prevent something very bad from happening, or alleviate someone's dire plight, by making only a slight (or even moderate) sacrifice, then it would be wrong not to do so' (p. 224).

¹⁶ Kaspar Lippert-Rasmussen (1996) raises a similar worry about Kamm's idea of moral status, grounded in the value of persons: why would the status of persons be determined only by inviolability, as Kamm maintains, rather than by other considerations as for instance by their non-sacrificibility (Lippert-Rasmussen's term)?

duties: the duty not to harm persons and to respect their rights, and the duty to help them when they are threatened to be killed. The standard examples for the paradox of deontology all illustrate just such a conflict.

But the ACR-involving view of rights does not solve this conflict, because it does not address it. It focuses on the conflict between not violating someone's right vs preventing rights violations of the same kind by others, but it is silent on the conflict between not violating someone's right and helping others. It may be taken (by some) to entail that whenever there is a conflict between the duty not to violate a person's rights and the duty to help, the duty not to violate the right wins out. But the ACR view of rights does not provide any support for this conclusion. (The duty to help simply does not come within its purview.) And surely it cannot draw on the value of human beings for support either.

The not-violating-rights vs helping conflict cannot be brushed aside by blaming it on the spectre of consequentialism. It is not wedded to consequentialism. The problem it poses is not due to any assumptions about minimizing rights violations, or maximizing the good. It arises because the respect for value requires both not destroying and protecting what is of value.

A number of solutions to the paradox of deontology have been proposed over the years. Deontological principles such as the Doctrine of Double Effect (DDE) or the Doctrine of Doing and Allowing (DDA) may explain why not violating a right takes precedence over preventing its violation, and why not harming a person is more important than not helping her—in both cases irrespective of the numbers involved. But of course these explanations are only as plausible as the proposed principles themselves. More recently, Frances Kamm and T. M. Scanlon have offered different explanations, which deserve to be explored further.¹⁷

But I am not going to discuss any of them here. I want to approach the Paradox of Deontology by taking the discussion away from the *all things considered* level of principles, and focus instead on understanding the *pro tanto* reasons that are involved. I hope that doing so will help with better understanding the nature of the conflicts.

In the following section I want to take the problem to a different area: the explanation of the duty to keep one's promises. My reason to shift

¹⁷ Kamm (2007); Scanlon (2008).

the focus at this point is that I want to clarify whether the problem is really with the structure of deontological restrictions—whether there actually is an ‘air of paradox’ to them—or whether the problem is of a different kind altogether.

II. DEONTOLOGICAL RESTRICTIONS AND PROMISE-KEEPING

The paradox of deontology supposedly arises with any agent-centred restriction. It is not only the restrictions on, or reasons against, killing which are thought of in this way, but also the reasons against promise-breaking, deceiving others, or torturing them (and possibly quite a few others).

Promising gives rise to agent-centred restrictions too. That is, assume that if agent A_1 keeps her promise to P_1 , then five other agents $A_2 \dots A_6$, will each break their promises to $P_2 \dots P_6$. It would be wrong for A_1 to break her promise even in this situation.¹⁸ Can we explain why this is so?

I will start with the reasons for keeping one’s promises generally, and then explain A_1 ’s reasons in a situation in which her promise-breaking would lead to a better outcome in terms of promises being kept.¹⁹

II. 1. Some general remarks on promises and reasons

The most striking feature of promising is that by giving a promise one makes it the case that one is under an obligation to keep it (provided the promise is valid). It does not follow, however, that we can conjure up obligations by mumbling ‘I promise’. That there is an act of promising is only one condition of a valid promise.²⁰ There also has to be a promisee who accepts the promise, and the promise must not be forced or immoral in content. All these are only necessary conditions for giving a valid promise, and I do not aspire to providing sufficient ones here.

¹⁸ As will become clear below, this is not true as it stands—it will need much qualification. But let it be our starting point.

¹⁹ I am aware that the view of promising that I will set out in this paper is controversial, and that I cannot adequately defend it here. I hope that it will not be too controversial to warrant discussion of the conclusions that I shall draw from it.

²⁰ I do not assume that promising actually requires the speech-act ‘I promise’. Depending on context, many different acts can constitute the giving of a promise. Similarly—and perhaps even more obviously—accepting a promise does not require the speech-act ‘I accept’.

The ability to make valid promises is a power of rational agents to change their normative situation.²¹ By promising to ϕ you put yourself under an obligation to ϕ , and you give the person to whom you make the promise the right to release you from that obligation. Thus a valid promise creates both an obligation to keep the promise and a right of the promisee to release the promisor.²²

Having this power is of value to agents in a number of ways. Two of them seem particularly salient to me. (1) The power to make valid promises enables an agent to bind herself to a certain course of action, and furthermore (2) it enables her to structure her relations with others in a certain way—to empower another person, for instance, by giving her some control over the situation that she would not otherwise have.²³ The reason why promises are binding, in general, is that it is of value to agents to have the power to bind themselves and to structure their relations to others in the way promising allows them to do. The reason to keep any particular promise is, however, simply that one promised. Having undertaken the obligation to ϕ I will (other things being equal) fail to act as I ought to do, if I do not ϕ . I will call this the *content-independent reason* for keeping a promise. It exists because agents have the power to undertake voluntary obligations, and they have this power because having it is of value in the two ways sketched above.

²¹ The explanation of promises that I propose owes a lot to Joseph Raz's (1977) account.

²² Pace Shiffrin (2008), it does not—I think—involve a 'transfer of rights'. For some reasons why not, see footnote 23.

²³ Seana Shiffrin (2008) very persuasively explains how promising can do the latter. I agree in part with Shiffrin's account, but not with her claim that by promising the promisor *transfers* the *right to decide* on the matter over which the promise ranges to the promisee. As stated above, I think the promisor gives the promisee *the right to release* her from the obligation she has undertaken. The right to decide what to do (if there is such a right at all) remains with the promisor. Two reasons for rejecting the rights-transfer view are as follows. (1) The reason for keeping one's promise can, on occasions, be defeated if, say, I promised to meet you for lunch but another person urgently needs my help. If by promising to meet you I had transferred the right to you to decide where I am going to be at 12 noon tomorrow, I would not have the right to decide that instead of meeting you I ought to help the person in need. But that is obviously false. The right I give you is different from the right to decide what I do. (2) But by promising I also do something else (and more) than giving you the right to decide: if I promise you to be there for lunch, I am under an obligation to be there. It is not that once I promise, you decide—as if everything was still up for grabs—but I have now (other things being equal) a sufficient reason to show up. So, in a sense the issue is settled. Hence, promising is centrally the undertaking of an obligation and not a rights-transfer, even though the promisee is given a right as well—the right to release the promisor.

If giving a valid promise *is* undertaking an obligation, there is no further question why there is an obligation to keep one's promises, and I assume that we have *pro tanto* reasons to comply with our obligations. The only question is why it is that we have the power to bind ourselves in this way—but that can be explained by its value.²⁴ The value of having the power to bind oneself also explains the exceptions: why, for example, immoral or coerced promises are not valid and therefore not binding. Coerced promises are invalid because they are not instances of an agent binding herself, while immoral promises are invalid because there is no value in having the power to create an obligation to act immorally.²⁵

²⁴ Norbert Anwander and Jan Gertken raised an important worry about this approach. Doesn't it involve a rather objectionable kind of boot-strapping? Perhaps a person can have an obligation to act in a certain way because it would be good if she so acted. But can she also have an obligation because it would be good if she had such an obligation? That, of course, is not what I claimed—but perhaps it brings home vividly why there may be some discomfort. My claim is that persons have the power to undertake obligations if and only if it is good that they have this power. I cannot fully defend this claim here. But let me make an attempt to dispel the worry. A normative power is a person's ability to change the normative situation (to make it the case that either she or another person has a reason that they would not have had otherwise). Provided certain conditions are satisfied, in exercising her power the person 'creates' a reason. But there are other ways of intentionally changing one's normative situation—not by exercising normative powers. Take an example. By intentionally hurting myself I can 'create' a reason to be taken to hospital. This reason can perhaps be explained in terms of value: if a person is hurt, there is a reason to take her to hospital, because it is good that she will be treated. I can intentionally make it the case that the antecedent is satisfied, thereby creating a reason. The explanation of normative powers is different: Here the idea is that it is good that a person can, by promising, create an obligation. Thus the claim is *not*, that if a person promises to ϕ , she has an obligation to ϕ , because it is good that she ϕ s, nor that she has such an obligation because it is good that she has that obligation. This is different from the hospital example, because there it is not true that there is a reason to be taken to hospital because it is good that by hurting myself I make it the case that I am entitled to a hospital bed—it is good that if I need hospital treatment I am entitled to it. That I can make it the case that I need such treatment has no value. Normative powers can be explained in terms of values, but not in the way in which other reasons may be so explained. While values play different roles when it comes to explaining reasons and explaining normative powers respectively, in neither case is there—as far as I can see—an objectionable kind of boot-strapping involved. The question is really whether this value-based explanation of normative powers is the right one. Norbert Anwander (2008) doubts this, but I do not have the space to discuss his view here. The understanding of normative powers on which I rely has been advanced by H. L. A. Hart (1994), pp. 43–4 (et passim.), and developed by Joseph Raz (1977). For a more recent defence of a similar view see David Owens (2006, and forthcoming).

²⁵ This may raise the question why silly promises (which I will discuss in more detail below) are valid when immoral promises are not. Is there any value in having the power to create obligations to do something which is not worthwhile doing? I believe there is. Remember that one of the points of promising is being able to structure one's relationships—to empower

There are, however, often *additional reasons*—reasons in addition to the content-independent ones—for keeping one's promises.²⁶ The promisee may expect me to act as promised, and I will let her down if I do not keep my promise; or she may have made plans relying on my acting as promised, and would be inconvenienced and disappointed if I did not keep my promise; or my acting as promised may simply be important (to the promisee or otherwise) not because it affects her relationship to me, or her plans, or her emotions, but because I promised to do something which is actually important and worthwhile.

But such additional reasons are only contingently related to promising. They are not essential because all additional reasons could exist whether or not there is a promise,²⁷ and the promise will be binding whether or not there are those reasons. Promises bind even when they are foolish, or when the promisee does not care at all whether or not the promisor keeps her promise, or when there is nothing to be said for acting as promised (because the promise is silly, for instance).

The content-independent reason to do as promised is independent of the additional reasons. Even if keeping my promise is not worthwhile I have a reason to do as I promised. Whatever it is that I promised (provided the promise is valid) there is a reason to act as promised, because by promising I undertook an obligation to do so. If I promised to get you a silly hat I have a reason for getting it, whether or not you look forward to my offering, or you would cherish it if I got it for you, or you would be disappointed if I did not. Even if you would not mind foregoing the promised treat, provided that I promised and that you accepted my promise, I am under an obligation to get it for

another person, but thereby perhaps also to bond with her. Silly promises can satisfy this role as much as any other, and therefore they are valid. Tom Hurka suggested to me that immoral promises are invalid because an agent cannot transfer the right to decide whether she should act immorally to another person, since she does not have this right to begin with. Silly promises are not affected by this concern. But this explanation depends on the rights-transfer view of promises which I reject (see footnote 23).

²⁶ Scanlon (1998) seems to assume that the obligation to keep one's promises has to be fully explained in terms of what I call 'additional reasons'.

²⁷ That is, there may be a reason to do something because someone else relies on my doing it (assuming that I know this), even if I did not promise to do so. This is so in cases where I made my intention to do something known to another person, or when I have firm habits of doing certain things so that others form expectations based on those habits. (For a more thorough discussion, see David Owens (2006).)

you. You may have accepted my promise only to humour me or not to offend me. Be that as it may. There is a reason to keep my promise which is necessarily related to promise-keeping. It comes into existence simply by promising, and because promising *is* the act of undertaking a voluntary obligation.

While the content-independent reason to keep one's promise—to comply with the obligation one has undertaken—is a particularly stringent reason (as are all reasons to comply with an obligation), it is by no means absolute.²⁸ Sometimes, breaking one's promise will be the right thing to do if the keeping of the promise conflicts with some more important duty or with a more important reason.

II. 2. The deontological restriction on promising

Let me now take this back to the deontological restriction. Isn't the case it describes perhaps a case in point where the reason to keep the promise is defeated? Ought I to break my promise if I can thereby make Anton (and his four companions) keep theirs? Like myself, by promising Anton undertook a voluntary obligation to act as promised. If he breaks his promise he fails to comply with an obligation that he imposed upon himself. Is there then a reason for me to see to it that he keeps his promise because he would be failing to comply with his obligation if he did not? Are there other reasons for me to bring it about that Anton keeps his promise?

Let me begin with the additional reasons for promise-keeping. Roughly, they fall into four categories:²⁹

²⁸ The reasons to comply with obligations are generally seen as particularly stringent. I believe that this is right, provided we use 'stringent' in a somewhat technical sense; it should *not* be understood as denoting 'important'. I mean by 'stringent' only that there is a particular structure of reasons that accompanies obligations. The reasons to comply with an obligation or a duty are special in that certain kinds of reasons that would count for or against acting in the required way (if there was no obligation) are simply irrelevant. Reasons from 'inclination', as Kant may have put it, are an obvious case in point. It does not matter whether I feel like getting you the silly hat once I have promised you one. While 'not feeling like doing something' may at times be or indicate a reason for not doing it, it does not when an obligation is involved. It is not that my inclination is 'outweighed' by the reason to keep my promise—it is simply irrelevant. I take it that this is the sense in which reasons to comply with one's obligations are particularly stringent, and different from reasons that do not involve obligations. The reasons are stringent in this sense, even if the obligation is petty (as in the case of the silly promise).

²⁹ This is really only a very rough classification. The categories overlap, and, if you prefer, you could regard all of them as outcome-related—distinguishing the outcomes by their contents. Nothing much turns on this.

(a) Outcome related:

- It is (independently of expectations, and so on) important that the promisor acts as promised.

(b) Promisee related:

- The promisee may expect the promisor to act as promised—so the promisor would let down the promisee if she did not keep her promise.
- The promisee may have made plans assuming that the promisor will act as promised, and would be inconvenienced if she did not.

(c) Relationship related:

- It may undermine trust between the promisee and the promisor if the promisor breaks her promise.

(d) Promisor related:

- A promisor may lose her ability to give valid promises if she breaks them frequently. (Her promises would no longer be accepted.)

Let us assume that Anton's promise to Paul is about something important and worthwhile; that Paul relies on Anton, and Anton would let Paul down badly if he would not keep his promise. Are these additional reasons reasons for me—the bystander—to see to it that Anton keeps his promise? Assume that Anton promised Paul to give him a lift to London to see his mother who has fallen ill (and there is no easy way to travel there fast otherwise). How does Anton's promise affect me? If Anton is about to break his promise, then as far as Paul's getting to London is concerned, I could give him a lift instead. I need not see to it that Anton keeps his promise, but only that the outcome will be secured: that Paul gets to London in good time. The reason to secure the outcome—while possibly being a reason for a mere bystander—is not *ipso facto* a reason to violate the agent-centred restriction. In particular, it is not a reason to violate the restriction, because doing so would prevent a greater number of violations of the restriction. It may, on some particular occasion, be a reason to break my own promise nonetheless. Perhaps I promised to see a friend for dinner, and if I take Paul to London I will not be able to keep this promise. But it would not be a breaking of a promise in order to minimize promise-breaking: as far as

the outcome-related reasons are concerned, it does not matter whether the outcome is being brought about by Anton keeping his promise, or in some other way.

There is also Paul's expectation and the possible disappointment. Should I break a promise of mine if that is what is needed to make Anton keep his promise, thereby avoiding these kinds of bad consequence? As opposed to outcome-related reasons, the only way in which Paul's disappointment could be avoided is presumably by Anton keeping his promise. Perhaps I could persuade Anton to keep his promise. Ought I to do so? The disruption of Paul's plans could provide a reason to bystanders, but his feeling of disappointment may be no reason for a bystander at all (at least not in general, special relationships apart) simply because his feelings would be entirely appropriate, given that Anton is about to break his promise. And it seems to me that there are no reasons to prevent a person from having appropriate emotions, even if experiencing them is not particularly pleasant.

But what about the relationship-related reasons? Should I try to see to it that Anton keeps his promise to make sure that there will be no friction in their relationship? As in the emotions case, frictions in people's relationships, when they adequately reflect the reasons that are relevant, are not *per se* regrettable. They may simply be appropriate.

Finally, there are the promisor-related reasons—reasons to make Anton keep his promise, because if he does not he would not be trusted in the future (by Paul, and perhaps even by others). Imagine that Anton is a friend of mine, or my child, and that he often does not keep his promises. He runs the risk that no one will accept his promises any more if he breaks this one too. Perhaps I worry about him in that regard. In that case, I may have a reason to do something to make him keep his promise—but whether 'something' would be the breaking of a promise of mine is again contingent, and perhaps unlikely, since my concern for Anton's trustworthiness may lose its credibility if I am readily willing to undermine my own.

Thus the only (additional) reasons that can easily be reasons for a bystander are the outcome-related reasons. All others could occasionally be reasons for a bystander, but we need to explain why they are when they are. They are not *ipso facto* reasons for everyone. And furthermore, none of the reasons for a bystander is obviously a reason to break her promise, in order to make Anton keep his. The outcome-

related reasons are reasons to do *something* to secure the outcome. The promisee-related reasons are reasons to do *something* to prevent the disruption of Paul's plans, his disappointment and his being let down, and so on.

Let us assume next that the promise is a silly promise and there aren't any additional reasons for keeping it (i.e., the outcome is not worthwhile, the promisee does not care about the promise being kept, and so on). In that case, as far as the additional reasons are concerned there obviously is no reason for a bystander to see to it that the promise will be kept—simply because there aren't any.

What about the content-independent reason for promise-keeping then? After all, this is the only reason that is non-contingently related to the promise. Do I have a reason to bring it about that Anton complies with this reason (all additional reasons apart)? There is an explanation why *Anton* should comply with his reason: he put himself under an obligation to act in a certain way. But perhaps all the reasons that can be shared, and where a bystander's effort to make someone keep his promise could make sense, are exhausted by the additional reasons that may or may not be present.

I think that this is in fact so. To make it more plausible, consider the following (Kantian) approach. If Anton does not keep his promise, one way of explaining what is objectionable here is that his will is not as it ought to be. He is not motivated to act as a moral person would be, or—as Kantians might express it—he is lacking in good will. Perhaps in some sense the world would be a better place if it had only people of good will in it. Would I have a reason to do something that makes it happen that others act out of good will, if I can do so? And is not that perhaps the same as a reason to make Anton comply with the content-independent reason to keep his promise? It would, after all, not depend on the particular content of the promise, or the outcome-related value of keeping it. Let me take these questions in turn. First, is there a reason for making it the case that others act out of good will? And second, if there were such a reason would it (in the promise-keeping case) be the same as the reason to make them comply with the content-independent reason to keep one's promises?

Relating to the first question: we can tempt others, thereby undermining their resolve and bringing it about that they fail to act out of

good will.³⁰ Doing this will often be wrong. If so, isn't it then perhaps also true that we should do what we can to strengthen another person's will (not only abstain from undermining her resolve) and make her comply with her obligations? As Kant sees it, whether an action is done out of good will makes all the difference to its moral worth, but not, however, to its permissibility.³¹ Should we neglect our own obligations to help others to be morally worthy agents and to act in morally worthy ways? This would not be a reason to see to it that the other person complies with her obligation, but more narrowly a reason to see to it that she does so for the right reasons.

We may have reasons not to undermine the general capacity of others to comply with their reasons, and to support them in acting for the right reasons. But ultimately an agent cannot be made to act for specific reasons. She can be encouraged (or discouraged), influenced, and supported. But it is not within the reach of my agency to bring it about that you act for the right reasons. Therefore, it cannot be the direct goal of my action to make this happen. The most likely way to at least influence the reasons for which others act is by reasoning with them. But we do not always have reason to do this. There are requirements of respect which draw the boundaries between mere meddling and good advice. Thus there is no reason to make others act out of good will, because another person cannot bring this about directly;³² and furthermore, there are limits to trying to influence others. Sometimes we have reason to do so, but not in general. If Anton is a mere stranger I still have reason to *hope* that he will turn out a person of good will, but it does not follow that I have reasons to try to do something about it.³³

Our second question was: should I violate my own obligation if this makes it more likely that you comply with yours—that you comply

³⁰ I am grateful to Michael Ridge for pressing me on this point.

³¹ Kant distinguishes between actions that are 'pflichtwidrig' ('contrary to duty') and actions that are 'pflichtmäßig' ('in accordance with duty'). If an action is 'pflichtmäßig', it can be done 'aus Pflicht' ('from duty') or out of self-love (GMS 397). An action has moral worth only if it is done from duty. Julia Markovits has argued recently that moral worth depends on the reasons for which an action is done: only if it is done for the right reasons does it have moral worth. Markovits (2010) steers away (rightly I think) from Kant's focus on duty.

³² I assume here that we do not have reasons to do what is conceptually impossible.

³³ I am grateful to Alex Voorhoeve for suggesting to me the distinction between reasons to hope and reasons to do.

with the content-independent reason for keeping your promise? The content-independent reason is a reason to act as you promised. It is not a reason to act, for the right reasons, as you promised. (Imagine that I promised you I would be in my office at a particular time, but I completely forgot about it. However, I am in my office anyway. I did keep my promise; I did comply with the content-independent reason; but I did not act as promised for the right reasons.) We are now considering whether I should break a promise of mine if by doing so I could make it more likely that you will keep your promise for the right reason (accepting that this is not under my direct control, but I may be able to influence you). Thus the aim of my action is not to make you comply with the content-independent reason, but to make you act in a morally worthy way. Therefore, it is not the content-independent reason that gives me a reason to break my promise, but a quite different concern anyway. If we furthermore accept that my reasons to improve the moral worthiness of others are quite limited (beyond in general trying not to undermine their capacity of acting for the right reasons), and have to be restrained by respect for the independence of autonomous agents, the Kantian approach does not provide us with a reason to make others comply with the content-independent reason to keep their promises. Perhaps it comes closest to explaining a mere concern for others complying with their obligations, independently of outcomes. But since it is not possible to make others act in morally worthy ways, nor always permissible to attempt to influence them, and since the concern for moral worth is different from the content-independent reason for promise-keeping anyway, the Kantian approach does not establish a reason to make others comply with the content-independent reason.

But if I do not even have a reason to make others act in morally worthy ways (as far as that is possible), how could I have a reason to make them comply with the content-independent reason for promise-keeping when nothing is at stake (when no additional reasons are present)? The only thing that my interference could bring about in this case is that a promise will be kept, where there is nothing to be said for the promise-keeping from the perspective of any of the persons involved. To claim that I have such a reason you would have to argue that a world with fewer promises broken in it is, in that regard, a better

world.³⁴ I am not sure that I even understand this claim, and I certainly see no reason to accept it.

If this is right, then it would not be permissible for me to break my promise when this would lead to fewer occurrences of promise breakings, since in doing so I violate the obligation to keep my promise for no reason at all.

In general, while you have a reason to keep your promises, I do not have a reason to break mine so that you keep yours (additional reasons apart). But even considering the additional reasons, this is still true for the most part, because the additional reasons are rarely reasons to see to it that the other person keeps her promise. They are mostly reasons to secure the outcome (when doing so is worthwhile).

But of course, there can be cases where the additional reasons would count in favour of breaking my promise to make you keep yours (where your keeping your promise is the only way to secure the outcome). But even in those cases, the reason why the obligation is defeated is not that more promises are being kept (or that fewer violations occur): that fewer violations occur is no reason at all to break a promise.

Thus, the obligation to keep ones promises turns out to be subject to ACRs in a quite natural way. It follows from our understanding of the value of promising that a bystander cannot realize its value. Its value lies in the power to bind oneself. Another person can perhaps assist with realizing whatever particular value the keeping of a particular promise may have, but having the power to bind oneself does not require assistance—and it would not be enhanced or protected by seeing to it that fewer promises will be broken.

II. 3. Agent-relative and personal reasons

Yet the content-independent reason is not simply an *agent-relative reason* (even though it is that too). After all, it has often been pointed out that agent-relative reasons give rise to reasons for others to

³⁴ Holly Smith suggested to me that I ignore the value of promise-keeping. In one sense this is quite right. I do not believe that there is value in promises being kept (unless there is value in the outcome of acting as promised, or some other additional reasons obtain). But let me remind you that I explained the reasons that we have for keeping our promises in terms of the value of having the power to promise. This is the point where value enters into the explanation, not at the stage of 'promises being kept'. But this is also why the value that explains our reasons for promise-keeping cannot be realized by a bystander.

‘promote’ the agent’s compliance with that reason.³⁵ The agent-relative reason of a mother to care for her child, for instance, may well give others reasons to support her and help her in her efforts to do so. That is why agent-relativity in itself cannot help us to solve the paradox of deontology.³⁶ The content-independent reason to keep one’s promise does not provide others with any reasons to ‘promote’ one’s keeping of the promise. Let me call reasons of this kind *personal reasons*, and understand ‘personal’ as indicating that there is a reason which is a reason only for the agent, and that its presence does not give any reasons to others.

If the content-independent reason to keep one’s promise is a *personal reason* in this sense, then it is not so very puzzling why the duty to keep one’s promises takes the form of an agent-centred restriction. There simply is no reason to break one’s promises to make another person keep hers, and violating one’s own obligation for no reason whatsoever would be impermissible.

Even the case in which I have reason to bring it about that you keep your promise by breaking a promise of mine is not a case of violating the agent-centred restriction. It rather is a case of breaking one’s promise, because doing so prevents some bad consequences when preventing them is more important than keeping the promise (and cannot be achieved in any other way).

Kamm raises an objection³⁷ against attempts to explain deontological restrictions by appealing to agent-relative reasons which may apply to personal reasons as well. An agent should not violate a deontological restriction to prevent a greater number of violations of that restriction—not even if the violation is necessary to keep *herself* from violating the restriction. Those who focus on agent-relativity cannot explain why this

³⁵ Nagel (1970), Parfit (1979), Sen (1982), and Korsgaard (1993) all argue (in different ways) that an agent-relative reason may well give rise to reasons for others to support the agent in her compliance with that reason. (Korsgaard rejects the label ‘agent-relativity’—not least because we can ‘share’ even those reasons which are based on personal pursuits.)

³⁶ As opposed to others, McNaughton and Rawling (2006) seem to think that the agent-relativity of a restriction or a reason is sufficient for establishing deontological restrictions. This seems doubtful to me, because agent-relative reasons may well give reasons to others: agent-relativity allows that A_1 has a reason to promote $A_2 \dots A_6$ ’s compliance with their agent-relative reasons not to violate the rights of $P_2 \dots P_6$ —which brings back the paradox of deontology.

³⁷ Kamm (1996), chap. 9. Kamm first discusses the issue in Kamm (1984).

is so, and personal reasons fare no better.³⁸ Since the agent's reasons for not violating the restriction once, as well as not violating it many times, are all of the same kind, it seems that—if numbers count—it would be better if she violated the restriction only once rather than many times. Kamm suggests that this is the wrong answer, and the agent-relativity of the reasons involved cannot explain why.

But is it the wrong answer? First, assume that I have given three promises, all equally important; that is, as far as additional reasons for keeping them are concerned they are all on a par, but I can keep two of them only if I break one. It seems to me that in this case I have most reason to break the one promise in order to make sure that I keep the two. At any rate, the question seems to be whether, in a case where all the options are on a par but an agent cannot comply with all her reasons, she should comply with the greater number of her reasons. The issue becomes one of aggregation. Secondly, assume that if I break a promise now, I can keep myself from breaking two promises that I am going to give in the future. In this scenario it seems possible that I can comply with all my reasons and not break any of my promises, if I keep the promise I have already given but abstain from making the future promises. If, for some reason, I cannot prevent myself from making those promises in the future, the case reverts back to the earlier one. Thus, as far as promise-keeping is concerned, Kamm's claim is false. But she had, of course, the case of killing in mind when she put forward her objection, and I will therefore return to it below.

Thus far there is at least one case where the agent-centred restriction does not look puzzling. The air of mystery that it is supposed to carry simply did not make an appearance in our discussion of promises (and we also did not have to appeal to deontological principles such as DDE or DDA to explain the restriction.)

But does this show that there is no puzzle after all? Or does it show that promising is a very special case, utterly different from other moral duties? I believe that promising is not just a special case, but it is nonetheless true that the killing cases are much more puzzling and complicated. The puzzle, as we can see now, is not due to the very

³⁸ Actually, the objection works better (if it works) as an objection to personal reasons for the reason already explained: agent-relative reasons fail to establish deontological restrictions anyway, because those reasons can provide reasons for others.

idea of agent-centred restrictions. That can be presented in a perfectly puzzle-free way. The puzzle pertains to the ethics of killing (or harming) others—or so I will argue.

III. DEONTOLOGICAL RESTRICTIONS AND KILLING

The reasons to keep one's promises only contingently combine content-independent and additional reasons; that is, there may be only the content-independent reason, and no additional reasons at all. Thus, it is possible that the obligation to keep one's promise is *personal* only (in the sense defined above): it gives no reasons to anyone other than the promisor to do anything about the promise-keeping.

The case against killing almost always combines two kinds of reason: reasons against violating another's right not to be killed, and reasons against killing³⁹ because it is harmful. There may be cases where the killing would not be harmful but still involve a rights violation,⁴⁰ as well as cases where the killing would not be harmful and not involve a rights violation (because the person has consented to being killed and thereby waived her right).⁴¹ Thus, the distinction between violating a person's right to life and harming her by killing her is not just an idle distinction. Furthermore, the harm of being killed is at stake in many cases where there is no rights violation, as when a trolley is hurtling towards potential victims. And, as we will see, the reasons for preventing harm are in some ways different from the reasons for preventing rights violations. Thus, while in the killing cases the rights violation almost always combines with harming, the two are distinct, as are the reasons for preventing them.

Distinguishing the two also helps to avoid a certain mistake. If the case is one of preventing a rights violation, then it involves other human agents (the violators of the rights) and their decisions. We may then conclude that avoiding rights violations is the violator's responsibility and no one else's. To borrow Thomson's memorable phrase: I do not have 'to make the villain's moral record better than it otherwise would

³⁹ I should probably speak more generally of the shortening or ending of a person's life.

⁴⁰ Cf. Lippert-Rasmussen (2007).

⁴¹ Perhaps some cases of euthanasia.

be'.⁴² This may be so. Yet there is a reason for me to interfere nonetheless—not in order to prevent the villain from besmirching her moral record or because I would be responsible for the killing if I fail to prevent it, but simply to help the victims. That I ought to help the victims is independent of the fact that the victims are threatened by a human agent who is responsible for the killings (that is, it is independent of the killing involving a rights violation). The victims of potential killers have a claim to being helped just as much as the victims of runaway trolleys or natural disasters. Some think that the claim to be helped is not a very strong one—but whatever it is, it is not (I think) further weakened by the fact that in the one case there is a killer involved.

I am going to argue that the ethics of killing is so vexing and puzzling because even if the reasons against the rights violation may take the form of ACRs, because they are personal reasons, the reasons against harming and the reasons for protecting others from being harmed do not.⁴³

But why think that reasons against violating someone's right are *personal reasons*? Here is a straightforward, intuitive way of making the case:

Rescuer's Decision. Imagine you find yourself in a situation where you could either save a person from being run over by a trolley or you could save someone from being killed by another person. (Only in the latter case does the killing involve a rights violation.) Let us assume that the death of the person involved is equally bad in each case: they are not going to suffer; they are of comparable age, health, and so on.

There seems no more reason, in this case, to save the one than the other. As long as you save one of them you have acted as you

⁴² Thomson (1985), p. 1414. Bernard Williams (1973) also argues that while an agent is responsible if she kills someone she is not responsible for someone else's killings. He is, however, careful not to draw the conclusion that there is therefore a deontological restriction on killing.

⁴³ The distinction can also help to show what is wrong with one of the defences of ACRs and deontology in general. It is sometimes said (even if not, to my knowledge, thoroughly argued) that while the killer is responsible for the killing, you are not. Therefore, you need not concern yourself with his killing. The mistake is that even though you are not responsible for the killing, the victim still has a claim on you to be helped—not because you are responsible but because you *can* help, and there is a general duty to aid.

ought to.⁴⁴ Hence, the rights violation does not give you a further or stronger reason to rescue the person who is about to be murdered. If so, then the reason not to violate another's right not to be killed may be a *personal reason* (in the sense explained above). The murderer has reason not to violate her victim's right, but you have a reason only to help the victim because of the harm of killing, and not in order to prevent a rights violation. That the killing occurs as a result of a rights violation does not make it worse.

Does *Rescuer's Decision* suffice to show that reasons against rights violations are *personal reasons*—that you must not violate anyone's rights, but that you need not intervene in order to prevent rights violations by another? This depends on whether the only or the best explanation for the rescuer's decision (assuming that it is correct) is that the reasons not to violate someone's rights are *personal reasons*. If we could establish that the reasons not to violate a person's rights are *personal reasons*, the conflict in the killing case would change its shape slightly. I began with the assumption that in ACR cases there are four kinds of reasons involved:

- 1) A_1 has a reason not to violate P_1 's right not to be killed.
- 2) A_1 has a reason not to harm P_1 .
- 3) A_1 has a reason to prevent $A_2 \dots A_6$'s rights violations.
- 4) A_1 has a reason to help $P_2 \dots P_6$.

Perhaps the rescuer's decision shows that we can dispense with (3). But (1) and (2) would still be in conflict with (4). Can we resolve the conflict in favour of (1) and (2), thereby perhaps vindicating the ACR? Would it be because (1) and (2) together are weightier than (4)? Would (1) on its own suffice to show that (4) has been defeated?

⁴⁴ Scanlon (1998) makes a similar point: '... even those who believe that there is a special prohibition against intentional killing do not think that one must, for *this* reason, prevent the murder rather than the accident. The idea that the murder is a much worse thing to have happen would seem to imply that one must try to prevent it, even if one has a slightly greater chance of being able to prevent the accidental death. But this seems wrong' (p. 83). Compare also the following passage from Nagel (1986): 'To use an example of T. M. Scanlon, if you have to choose between saving someone from being murdered and saving someone else from being killed in a similar manner accidentally, and you have no special relation to either of them, it seems that your choice should depend only on which one you are more likely to succeed in saving' (p. 178).

I will not try to answer this question here. But is it even the right question on which to fix? You may have been quite surprised to hear that there are no reasons of type (3). Especially in the political arena, and regarding certain kinds of rights, it seems that preventing a rights violation is a very strong reason—perhaps even the only reason—for a bystander to intervene. Or is it?

Take freedom of expression. Assume that every person has a right of freedom of expression against all other people,⁴⁵ and that its justification lies in the value to the right-holders of having freedom of expression (the condition in which they have unimpeded access to the means of expression). Call this the *ground of the right*. It seems that there is a reason to see to it that speakers have the freedom to exercise their right—that is, that they enjoy the value which is the ground of the right: they can make their voices be heard publicly. Whether or not there is an outcome-related reason for *exercising* their right to freedom of expression is here once again a contingent matter. Perhaps their ability to use the freedom sensibly or wisely is limited or non-existent, so that exercising their right would not contribute anything of value. That is, we should distinguish between the value of exercising one's right to free speech, and the value of having the freedom to do so. My suggestion is that only the latter is the ground of the right.

In general, in societies like ours, the right is also a right against the government, and the responsibility of securing freedom of expression will lie with law enforcement agents. But imagine that there is no policeman anywhere in reach, and that someone's right to free speech is about to be violated. In that case there may well be a reason for anyone who can do so to act in a way that will prevent the violation of the right. Does this show that there are type (3) reasons after all? It does not, because it may well be consistent with the personal nature of the right. The reason that bystanders have in such cases would be the benefit for the person concerned of having freedom of expression. That, rather than the prevention of the rights violation, may require intervention by bystanders. It is an outcome-related reason, yet the outcome in question

⁴⁵ Alternatively, it could be a right against the government only. I assume here that it is both a right against the government and against other people. We do not have to settle this issue here. The important point for my discussion is only to show that understanding such a right (in either of these ways) does not involve type (3) reasons.

is not the value of the speech but the value of having the freedom. The reason for government protection and intervention is the same: it is securing the benefit of freedom to the person whose right is being violated that explains why policemen ought to stop the interference. The reason is not that they are preventing or minimizing rights violations.⁴⁶ We can explain why there is a reason (or, in the case of governments, a requirement) to intervene when someone's rights are violated without appealing to type (3) reasons.

It follows from this explanation that a bystander's reasons to intervene may have to take into account how the outcome—securing the freedom—can best be realized. That is, if by not intervening in one case but rather in another, it is more likely that the freedom which is the ground of the right will be secured (perhaps for a greater number of people)—that is what the bystander ought to do. Thus her reasons are different from the agent's reasons against violating the right: the violator should abstain from violating the right even if by doing so she can prevent a greater number of rights violations, because each person has a right against her to free speech.⁴⁷ But the person whose right is being violated does not have a right against the bystander to prevent the violation: there just is a reason to help for everyone (including the bystander) to secure the ground of the right (the value that the right protects). If so, the ability to prevent a greater number of rights violations may count from a bystander's perspective in a way in which it does not count from the perspective of the person who is about to violate a right.⁴⁸ But it is not the minimizing of the rights violation—even in the bystander's case—which furnishes the reason, but the ability to

⁴⁶ Possibly, the police (as well as all others) sometimes have a reason to prevent certain rights violation in order to publicly assert the value of the right. But here again the reason would not be preventing the rights violation as such.

⁴⁷ This is actually oversimplified. It follows from my analysis that the situation is more complicated. If considering whether to violate someone's right in order to prevent a greater number of rights violations, the way the reasons lie (if I am right) is this. An agent, about to violate someone's right has (1) a personal reason not to violate the right and (2) an outcome-related reason not to violate her right (where the outcome is securing the freedom). The bystander has (3) an outcome-related reason to protect other people's rights. There just is *no* reason to prevent a rights violation in order to minimize rights violations. In the ACR cases, however, the agent has type (1) and (2) reasons, conflicting with type (3), and—as in the killing case—whether she ought to violate one person's right in order to help a greater number will depend on how to resolve conflicts of this kind.

⁴⁸ I am grateful to Peter Railton for pressing me on this point.

secure the outcome (everyone's freedom to speak up in public). Securing this outcome is presumably not simply a function of keeping the numbers of rights violations down, but may have to do with the severity (and perhaps other features) of the rights violations.

How far does this get us towards explaining the deontological restriction on killing? It allows us to say the following. There is no reason for violating a person's right not to be killed when doing so would prevent a greater number of rights violations—simply because there is no reason to prevent rights violations *per se*. There is only a reason not to commit them—and that is, *after all, what the agent-centred restriction was all about*. So if the argument above is correct, we have established the agent-centred restriction not only for promises but also for killing. And once again, there does not seem to be an air of paradox, nor a puzzle. It is explained by the personal nature of the reasons against rights violations.

Let me return to Kamm's objection against this kind of explanation. According to Kamm, an agent should not violate someone's right not to be killed, even if by doing so she could prevent *herself* from a greater number of rights violations of the same kind. The personal nature of reasons against rights violations cannot explain this, because in this case all the reasons involved are personal. In the promising case I rejected the claim on which Kamm's objection rests: the claim that it would be wrong of an agent to violate a deontological restriction in order to prevent herself from a greater number of violations of the same restriction. In the current case I am inclined to do the same. The options of my killing one in order to prevent myself from killing five, and abstaining from killing one and therefore killing five, look to me very much like: my killing either one or five, and in that case we may be allowed to aggregate. Killing one is wrong, but it is the lesser wrong if the only available alternative is killing five.⁴⁹ Therefore, Kamm's objection fails: it is not true that I should not kill one, even if doing so is necessary to prevent myself from killing five. Presumably, Kamm would contest the equivalence of the two formulations ('killing one in order to prevent myself from

⁴⁹ This presupposes that numbers matter. Thomson (2008) endorses the following principle that she takes from Foot (1967): '*Killing Five Vs. Killing One Principle*: A must not kill five if he can instead kill one.' She finds aggregation here no more problematic than in rescue cases (where it has been most discussed).

killing five', and 'killing either one or five')—claiming, perhaps, that there is (or can be) a crucial difference between simply killing a person and killing her in order to save others from being killed.⁵⁰ I fail to see that difference—and if there isn't any such difference, the crucial premise on which Kamm's objection rests—that an agent must not kill one, even to prevent herself from killing several—is false.

However, the reason why killing one would be preferable is not that it prevents a greater number of rights violations (by the agent). Remember, we said that there is no reason to prevent a rights violation *per se*. Thus, minimizing rights violations is not the issue. In Kamm's cases, the reasons involved are:

- i) A_1 has a reason not to violate P_1 's right not to be killed.
- ii) A_1 has a reason not to harm P_1 .
- iii) A_1 has reasons not to violate $P_2 \dots P_6$'s rights not to be killed.
- iv) A_1 has reasons not to harm $P_2 \dots P_6$.

My claim that the agent ought to violate the one person's right rather than the five persons' rights does not show that A_1 should minimize rights violations after all. Instead, it shows that when facing the option of violating one person's or many persons' rights, numbers make a difference: not the number of rights violations, but the number of people being harmed. But, of course, if A_1 decides to violate only P_1 's right, and not $P_2 \dots P_6$'s, she commits a wrong. She would be justified in doing so only if she had an excuse or a justification (if, say, she could not have prevented being in a situation in which she has to violate someone's right). Failing that, violating P_1 's right is only the lesser of two wrongs.

If it is true that there are no type (3) reasons, there is a deontological restriction on killing in this sense: an agent ought not to violate a person's right not to be killed in order to minimize the number of rights violations of the same kind. Yet, I suspect that this result does not deliver all that a deontologist would have hoped for. I have *not* shown that it would be wrong to kill one person in order to prevent another person from killing five. I have only shown that the minimizing of rights

⁵⁰ She offers, as an example, a case where 'an agent has set a bomb that will kill five people unless he himself now shoots one other person and places that person's body over the bomb' (1996, p. 242).

violation is not a reason to do so. But there are still type (4) reasons—reasons to help those in danger. And I am not convinced that reasons of this kind are particularly ‘weak’. Thus it may be that you ought to kill the one after all, to help the five.

In this regard the killing cases are different from the promising case. Regarding promises, we can perhaps say that, as a default assumption, you ought not to break your promise in order to make another person keep hers. But there is no such default in the killing cases. Consider the relevant reasons again. Promise-keeping comprises the following three kinds of reason:

- 5) A_1 has a content-independent reason to keep her own promise P_1 .
- 6) A_1 may (contingently) have additional reasons to keep her own promise to P_1 (outcome related; due to their relationship; to avoid becoming untrustworthy).
- 7) A_1 may (contingently) have reasons to secure the outcome of $A_2 \dots A_6$'s promises, or to prevent P_2 's disappointment, the loss of trust between P_2 and A_2 , or of A_2 's ability to give valid promises; and some of those reasons may be due to special relationships between A_1 and $A_2 \dots A_6$, or $P_2 \dots P_6$ respectively.

The ACR-interpretation of promise-keeping is correct because type (5) reasons are *personal reasons*: there is no reason to break one's promise just because doing so would make another person keep hers. There simply are no reasons (conflicting with (5)) of the form:

- 8) A_1 has a reason to see to it that A_2 keeps her promise because of the content-independent reason to keep the promise.

Type (7) reasons are not reasons to prevent the breaking of a promise. Their explanation has to do with different features that occasionally can accompany the breaking of a promise. Thus there isn't always and necessarily a conflict of reasons in the cases in which the ACR description applies (because there may be no type (7) reasons present at all, or if there are such reasons, they can be complied with without breaking one's promise). If there are type (7) reasons present, and complying with them requires that one breaks one's own promise, there is a conflict, and it has to be resolved by taking into account the relative importance of the promises and the other considerations. It is not, in those cases, guaranteed that a bystander could not have most reason to break her

promise in order to make another person keep his, but the explanation of this case would not appeal to minimizing the violations of the constraint.

The killing case is different, because there are, in addition to type (1) reasons, almost always reasons of types (2) and (4) involved. Can the conflict be resolved—in parallel to the promising case when there is a conflict—by appealing to the relative importance of the considerations involved? The answer depends on the answers to two other questions: (i) How are we to resolve a conflict between violating a person's right and helping others generally? (ii) Do numbers count (in conflicts of this kind)—might the helping be more important, if greater numbers are involved? I do not have an answer to either of these questions; and at any rate, dealing with them is for a different occasion.

IV. CONCLUSION

I have shown that there may be a type of reason—which I called *personal reasons*—that are similar to certain kinds of agent-relative reasons in some ways, but different in others. They are reasons not to act in certain ways which do not provide reasons for others to support the agent in complying with the reason. If there are such reasons, and if reasons for not violating deontological restrictions are of this kind, we have a ready explanation of why they give rise to agent-centred restrictions: there is no reason to *minimize* the violation of the restriction, because there simply is no reason at all to prevent others from violating the restriction. If this is correct, we now also have an explanation why the ACR-interpretation of rights is correct, and the MV-interpretation of rights ought to be rejected. If there are no reasons to minimize rights violations, an interpretation of rights which understood them as being protected only up to the point at which their violation would prevent a greater number of rights violations, would be completely unmotivated.

But this conclusion is more limited than it may at first appear. It shows only that if we focus on a certain kind of conflict—the conflict between keeping one's promise, and preventing another person from breaking hers (when there are only content-independent reasons involved), or the conflict between violating a person's right not to be killed and preventing someone else from violating such a right—the

answer is clear: there is a reason to keep one's own promise, but none to make another person keep hers. There is a reason not to violate a person's rights, but none to prevent the rights violations by others. But that does not give us a conclusive answer what to do, because sometimes (in the promising case) and almost always (in the killing case) there are other reasons involved, and they may well count in favour of violating the restriction.

But perhaps this is as it ought to be. Bernard Williams urged⁵¹ that we want to understand why certain decisions appear so vexing and difficult. The ethics of killing in particular is troubling and confusing—and a philosophical explanation should not assume otherwise.⁵²

REFERENCES

- Anwander, Norbert (2008). *Versprechen und Verpflichten*. Paderborn: mentis.
- Foot, Philippa (1967). 'The Problem of Abortion and the Doctrine of Double Effect', reprinted in her *Virtues and Vices*, pp. 19–32. Oxford: Clarendon Press, 2002.
- Hart, H. L. A. (1994). *The Concept of Law*, second edn. Oxford: Clarendon Press.
- Kamm, Frances (1984). 'Constraints and You', paper presented at the APA, Pacific Division Meeting.
- (1996). *Morality, Mortality*, vol. 2. Oxford: Oxford University Press.
- (2007). *Intricate Ethics*. Oxford: Oxford University Press.
- Kant, Immanuel (1785). *Grundlegung zur Metaphysik der Sitten*. Quoted from the Akademie Ausgabe, Vol. IV. Berlin, 1903.

⁵¹ Bernard Williams (1973), p. 99.

⁵² I presented earlier versions of this paper at the seminar of the Safra Center at Harvard, the Leeds CEM seminar, the philosophy colloquium at York University, the First Annual Workshop in Normative Ethics in Arizona, the practical philosophy colloquium at Humboldt University, Berlin, and the philosophy colloquium at Warwick University. My warmest thanks are extended to all these audiences for very helpful comments and stimulating discussions. In particular I would like to thank Keith Allen, Norbert Anwander, Arthur Applbaum, Anne Barnhill, Anne Burghard, Alix Cohen, Daniel Elstein, Jan Gertken, Stephen Holland, Tom Hurka, Wouter Kalf, Peter Lamarque, Gerald Lang, Paul Noordhof, Christian Piller, Peter Railton, Mark Reiff, Tanina Rostain, Thomas Schmidt, Mark Schroeder, Holly Smith, Mark Timmons, Pekka Väyrynen, and Alex Voorhoeve, for their comments which helped to improve the paper in many more ways than I can properly acknowledge. I would also like to thank Christian Miller and Michael Ridge for their comments and suggestions. Finally, I am especially grateful to Kasper Lippert-Rasmussen and Joseph Raz for extremely helpful written comments.

- Korsgaard, Christine (1993). 'The reasons we can share: An attack on the distinction between agent-relative and agent-neutral values', *Social Philosophy and Policy*, 10(1): 24–51.
- Lippert-Rasmussen, Kasper (1996). 'The Impermissibility of Minimizing Violations', *Philosophy and Public Affairs*, 25(4): 333–51.
- (2007). 'Why Killing Some People is More Seriously Wrong than Killing Others', *Ethics*, 117(4): 716–38.
- McNaughton, David, and Rawling, Piers (2006). 'Deontology', in *The Oxford Handbook of Ethical Theory*, ed. D. Copp, pp. 424–58. Oxford: Oxford University Press.
- Markovits, Julia (2010). 'Acting for the Right Reasons', *The Philosophical Review*, 119(2): 201–42.
- Nagel, Thomas (1970). 'Objective Reasons', in *The Possibility of Altruism*, pp. 90–8. Princeton: Princeton University Press.
- (1986). 'Autonomy and Deontology', in *The View from Nowhere*, pp. 164–88. New York: Oxford University Press.
- Narveson, Jan (2003). 'We Don't Owe Them a Thing!', *The Monist*, 86(3): 419–33.
- Nozick, Robert (1974). *Anarchy, State and Utopia*. New York: Basic Books.
- Owens, David (2006). 'A Simple Theory of Promising', *Philosophical Review*, 115(1): 51–77.
- (forthcoming). *Shaping the Normative Landscape*. Oxford: Oxford University Press.
- Parfit, Derek (1979). 'Is Common-Sense Morality Self-Defeating?', *Journal of Philosophy*, 76(10): 533–45.
- Raz, Joseph (1977). 'Promises and Obligations', in *Law, Morality and Society*, ed. Peter Hacker and Joseph Raz, pp. 210–28. Oxford: Oxford University Press.
- Scanlon, T. M. (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- (2008). *Moral Dimensions*. Cambridge, MA: Harvard University Press.
- Scheffler, Samuel (1982). 'The Defence of Agent-Centered Restrictions', in *The Rejection of Consequentialism*, second edn, pp. 80–114. Oxford: Clarendon Press, 1994.
- (1985). 'Agent-centered restrictions, rationality and the virtues', *Mind*, 94: 409–19; reprinted in his *The Rejection of Consequentialism*, second edn. Oxford: Clarendon Press, 1994.
- Sen, Amartya (1982). 'Rights and Agency', *Philosophy and Public Affairs*, 11(1): 3–39.
- Shiffrin, Seana (2008). 'Promising, Conventionalism and Intimate Relationships', *Philosophical Review*, 117: 481–524.
- Singer, Peter (1972). 'Famine, Affluence and Morality', *Philosophy and Public Affairs*, 1(1): 229–43.

- Thomson, Judith Jarvis (1985). 'The Trolley Problem', *The Yale Law Journal*, 94: 1395–415.
- (2008). 'Turning the Trolley', *Philosophy and Public Affairs*, 36(4): 359–74.
- Williams, Bernard (1973). 'A Critique of Utilitarianism', in J. J. C. Smart and B. Williams, *Utilitarianism For and Against*, pp. 77–150. Cambridge: Cambridge University Press, 1973.