Side constraints and the structure of commonsense ethics*

Theresa Lopez, Jennifer Zamzow, Michael Gill, and Shaun Nichols
University of Arizona

Introduction

In our everyday moral deliberations, we attend to two central types of considerations – outcomes and moral rules. How these considerations interrelate is central to the long-standing debate between deontologists and utilitarians. Is the weight we attach to moral rules reducible to their conduciveness to good outcomes (as many utilitarians claim)? Or do we take moral rules to be absolute constraints on action that normatively trump outcomes (as many deontologists claim)? Arguments over these issues characteristically appeal to commonsense intuitions about various cases. As a result, an important portion of the debate involves empirically tractable questions — questions that can be investigated by probing for people’s judgments in cases in which the two types of considerations come into conflict with one another.

In this paper we present a series of experiments that investigate how outcomes and rules enter into the deliberations of agents. Our findings point to an important feature of commonsense morality that many normative ethicists have overlooked: people are much more inclined to subject potential violations of moral rules to cost-benefit analyses when doing so will lead to fewer of their own violations (intra-agent scenarios) than when doing so will lead to fewer such violations in general (inter-agent scenarios).

That commonsense morality has this feature is not decisive proof that this feature ought to be part of our moral deliberations. But as we will see, some of moral philosophy’s most influential arguments for both utilitarianism and deontology rely on claims about commonsense morality that

* We’d like to thank Doug Portmore, Mark Timmons, and Jamie Dreier for helpful comments on an earlier draft of this paper.
fail to take this feature into account. We hope to show that those arguments — on both sides of the utilitarianism-deontology divide — are vitiated by this failure.

§1 Side Constraints and the Critique of Consequentialism

A cluster of familiar objections to act-utilitarianism maintains that act-utilitarianism is defective because it focuses exclusively on outcomes, neglecting other morally relevant considerations. Typically the objections take the form of thought experiments in which we find ourselves rebelling against the idea that we should simply seek to maximize utility. The magistrate and the mob\(^1\) case provides a representative example. Imagine that a crime is committed and the town’s citizens are outraged. If someone is not captured and punished, a vengeful mob will hit the streets. However, the town’s sheriff could make an innocent man pay for the crime in order to prevent the riots. Supposing that punishing the innocent man would lead to greater overall utility, it appears the act-utilitarian would require the sheriff to punish the man. But punishing the innocent man is clearly wrong.

Act-utilitarians have ways to respond to this objection, the most obvious of which is to emphasize that the calculations their theory involves are more complex and take more factors into account than the description of the case suggests (e.g. Hare 1981). Yet the critic pushes the point and argues that the case illustrates a deeper problem with utilitarianism: utilitarianism leaves out something fundamental. But what? One obvious candidate is consideration for the innocent man’s rights; the utilitarian calculation as described is inattentive to rights and their nonviolation. As Nozick (1974, 28) notes, this objection might be construed as claiming that utilitarianism suffers from an overly narrow conception of the good or the end state of affairs to be aimed at in action.

If the objection is just that act-utilitarianism is blind to rights, however, it admits of a consequentialist solution. For the consequentialist might maintain that the non-violation of rights is an outcome that we want to maximize. But Nozick maintains that amending the theory to include non-derivative consideration for rights and their nonviolation is not adequate to address the deep problem. For a theory may attend to the nonviolation of rights yet do so “in the wrong place and in the wrong manner” (28). A consequentialist view according to which violations of rights are to be

\(^1\) An early formulation of this objection appears in McCloskey (1957).
minimized would still require us to violate someone’s rights if doing so would minimize the total amount of rights violations in the society. In the magistrate and the mob case, if punishing the innocent man will prevent many others from having their rights violated, then we will be required to violate the innocent man’s rights by punishing him. Hence, Nozick maintains, even on a consequentialist theory that is attentive to rights violations, we might still be required to do something wrong.

The deep problem with utilitarianism, according to Nozick, is not in its conception of the good, but in its theoretical structure. Nozick proposes that rather than incorporate rights into the end state to be achieved, we ought to take them to be side constraints upon actions to be done (29). The constraints forbid certain types of actions irrespective of what benefits such actions might produce. Side constraints operate outside of the goals to be achieved by one’s actions and thus outside of cost-benefit analysis:

The side-constraint view forbids you to violate these moral constraints in the pursuit of your goals; whereas the view whose objective is to minimize the violation of these rights allows you to violate the rights (the constraints) in order to lessen their total violation in the society (29).

Nozick grounds the side-constraint view in the Kantian notion that individuals are inviolable and may not be treated as mere means. Nozick stops somewhat short of characterizing these constraints as absolute. He writes, “The question of whether these side constraints are absolute, or whether they may be violated in order to avoid catastrophic moral horror… is one I hope largely to avoid (30).” So Nozick leaves open whether violating a constraint may be permissible in some extreme circumstances (e.g., violating one person’s rights in order to save the lives of thousands). But Nozick doesn’t take this to undercut the idea that some moral considerations act as side constraints. On the contrary, he maintains, in the typical case it is unnecessary and inappropriate to perform cost-benefit analyses to figure out whether violating a person’s rights is the right thing to do all things considered. Side constraints are, if not entirely non-overrideable, overrideable only in

---

2 Nozick takes the strong intuitive plausibility of his assessment of the ‘magistrate and the mob’ case as evidence for the existence of side constraints, though he ultimately grounds side constraints in the Kantian idea that persons are separate and inviolable (34). So while the side constraint view does not ultimately depend on common sense intuitions, such intuitions are central to getting a grasp on the proper form of the side constraint view in the first place.
catastrophic situations. Nozick’s point is that cases like the magistrate and the mob indicate that many moral rules – such as those against lying, killing, and breaking promises – operate as side constraints on the pursuit of our aims, desires, and interests. Catastrophic situations notwithstanding, Nozick thinks that the side constraints operate outside of cost-benefit analysis and constrain what one may do in the service of one’s goals.³

§2 Resisting Side Constraints

Utilitarians, of course, reject Kantian or Nozickian side constraints. One important line of utilitarian argument is that such side constraints are incompatible with commonsense morality. Sidgwick makes such an argument by deploying Kant’s murderer at the door case: A friend pursued by a murderous villain seeks refuge in your home. His would-be murderer arrives at your door and asks whether your friend is there. Is it morally permissible for you to lie to the villain to save your friend? Kant says no, contending that telling the truth is “a human being’s duty to everyone, however great the disadvantage to him or to another that may result from it” (Kant 1797). Kant thus asserts a strict side-constraint view, forbidding any calculation about whether or not to lie. But Sidgwick plausibly contends that commonsense rejects Kant’s view in this case. Commonsense, according to Sidgwick, holds that while one may have a general right not to be lied to, such a right is forfeited or suspended in certain circumstances. As Sidgwick puts it, “[I]t seems strange if we may not lie, if lying will defend us better against a palpable invasion of our rights… Common Sense does not seem to prohibit this decisively” (Sidgwick 1874, 312).

Crucially, however, Sidgwick then goes on to claim that the commonsense intuition that it is acceptable to lie to the murderer at the door shows not merely that Kant’s side-constraint view is incorrect but that utilitarianism in general is correct. He writes, “[I]f the lawfulness of benevolent deception in any case be admitted, I do not see how we can decide when and how far it is admissible, except by considerations of expediency; that is, by weighing the gain of any particular deception against the imperilment of mutual confidence involved in all violations of truth” (Sidgwick 1874, 316). Kant’s blanket moral prohibition on calculating the effects of lying indeed

³ In this paper, we will simply grant Nozick the claim that catastrophe cases don’t provide an insuperable objection to a side-constraint account.
seems incorrect. And Sidgwick takes this to reflect a deep reason to adopt utilitarian calculation rather than deontological prohibitions.

Peter Singer and Joshua Greene are among the intellectual heirs of this argument for comprehensive utilitarian calculation, holding that correct moral judgments are those that hew to cost-benefit analysis of any and all of the results of an action. According to Singer and Greene, the problem with the deontological principles is precisely the fact that they are completely insensitive to costs and benefits. They fail to engage in any calculation at all, functioning instead as crude alarms that issue simplistic commands — “Don’t do it!” or “Must do it!” — in response to a single feature of what are in fact complex and nuanced situations (Greene 2008, 64; Singer 2005).

We thus seem forced into choosing between a side-constraint view that forbids calculation of effects and a utilitarian view that makes the rightness of the following of rules always contingent on comprehensive calculations of all effects. But in fact commonsense moral judgment might not fit well with either of those two options. As we will try to show, commonsense moral judgments are guided by a certain kind of constrained calculation that is captured neither by the notion of side constraints nor by the comprehensive utilitarian cost-benefit weighing of all possible effects.

§3 Confounds

As we’ve seen, Nozick maintains that utilitarianism has the wrong structure, and he maintains that the right way to think about normative ethics involves side constraints. Cases like magistrate and the mob motivate this side-constraint approach. However, this case is confounded. The case concerns whether an agent should commit a moral violation in order to forestall others from committing more extensive moral violations. That is, this case invites us to consider whether cost-benefit analysis is appropriate for inter-agent scenarios – scenarios in which one agent can commit a violation to prevent others from committing similar violations. So, the fact that we find it intuitively unacceptable might be because we take the rule to be a side constraint, or it might be because the calculation is across agents. People might be more inclined to engage in cost-benefit analyses when an agent is faced with committing a moral violation to avoid his own commission of more extensive moral violations. If such intra-agent scenarios still provoke a prevailing judgment that one should not commit the violation to minimize one’s own violations, this would suggest that the moral rules do operate as side constraints. However, if in such intra-agent cases people are more willing to say that one should commit a violation to preempt more extensive violations, this would suggest that people
are not so resistant to cost-benefit analyses after all. The question is of real significance. For if we do find that people are sensitive to calculations in intra-agent cases, then we cannot take the familiar intuitions about the magistrate and the mob case either as motivation for side constraints or as reason to think that the structure of commonsense ethics is best captured in terms of side constraints.

At the same time, the possibility that commonsense morality is especially sensitive to calculation in intra-personal cases could reveal a deep flaw in some of the arguments Sidgwick and his descendents deploy for utilitarianism. We may think it right to engage in a certain kind of moral calculation when trying to decide whether to lie or break a promise in intra-agent cases. But that doesn’t mean that people are committed to the significantly different, utilitarian kind of calculation, which asks us to engage in a comprehensive inter-agent weighing of all costs and benefits in every case. That is, people may reject the comprehensive cost-benefit analyses of utilitarianism, yet remain sensitive to cost-benefit calculation in more restricted domains. And if the commonsense application of moral rules does reflect a certain kind of calculative sensitivity in certain domains, then the rules cannot be dismissed as crude alarm bells.

§4 Experiments on Side Constraints

Nozick leaves open whether the side constraint view allows that rules may be overrideable in catastrophe scenarios. However, we wanted to explore whether people would allow cost-benefit reasoning to guide their judgment in intra-agent cases even when the stakes are not catastrophic. We have devised a number of paired cases to explore this hypothesis. The scenarios we used are not the familiar dramatic scenarios involving runaway trolleys and overzealous surgeons. We used more mundane examples to avoid worries that the dilemmas used in the standard dilemma literature are too far removed from the sorts of dilemmas people actually face. Our dilemmas involve situations like breaking promises and lying to one’s parents that we can expect to have a real world familiarity for most people. We constructed pairs of cases that are closely parallel except for one feature. In each pair, for one case – the inter-agent case – the agent can minimize moral violations in general, and for the other case – the intra-agent case – the agent can minimize his own moral violations.
**Experiment 1**

For our first experiment, we used two different kinds of scenarios, one involving providing support for one’s child and the other involving breaking a promise. For each scenario, half of the subjects were presented with the *inter-agent* case. In the promise-breaking case, subjects were given the following scenario:

Three people in Joe’s community, Mark, Bill, and Frank, are planning to move over a one-week period. Joe has promised Mark that he would help him move. Meanwhile David has promised Bill and Frank that he would help each of them move. When the time comes to help, it turns out that the only way for David to keep his promises to Bill and Frank is if Joe drives him, which would require Joe to break his promise to Mark.

They were then asked to indicate agreement (on a 7-point scale) with the following statement:

All things considered, Joe should break his promise to Mark, so that David can keep his promises to Bill and Frank.

In both of the *inter-agent* cases, subjects gave characteristically Nozickian responses. Just as in the magistrate and the mob, people tended to reject the idea that the agent should act so as to minimize violations *in general*. For instance, in the promise-breaking case, people tended to disagree with the claim that Joe should break his promise so that David can keep his.

Let’s turn now to the *intra-agent* versions. In the promise-breaking scenario, the *intra-agent* case went as follows:

Three people in Joe’s community, Mark, Bill, and Frank, are planning to move over a one-week period. Joe has promised each of them that he would help them move. When the time comes to do so, Joe realizes that he cannot keep his promises to all three. He can either keep his promise to Mark and break his promises to Bill and Frank, or he can keep his promises to Bill and Frank and break his promise to Mark.

The participants who received this question were asked to rate agreement with the statement:

All things considered, Joe should break his promise to Mark, so that he can keep his promises to Bill and Frank.
Here the responses were quite different. For the *intra-agent* scenarios, subjects tended to agree that Joe *should* break his one promise so that he can keep two other promises. The same pattern emerges for the child-support case.\(^4\)

**Experiment 2**

One might worry that our first two cases unfairly narrow the choice space. For the *only* thing to guide the choice in the intra-agent condition is the number of violations. Our results already suggest that people are not oblivious to calculations, as side-constraint theories are. But it would be more compelling if people were sensitive to calculations even when the choice is between violations of different kinds. As a result, we conducted a second study in which an agent is faced with committing a moral violation in order to preempt more extensive moral violations of a different sort. In particular, an agent has to decide whether to lie to his parents in order to avoid multiple

\(^4\) The statistical details for the promise-breaking case are as follows. The mean response in the inter-agent version is 2.37 (out of 7); the mean response in the intra-agent version is 4.89. The difference between the conditions is statistically significant (\(t (35) = 5.263, p<.001\)).

For the child-support cases, in the inter-agent condition, subjects received the following case:

Steve has one young child. His child lives in an area with a high cost of living. Steve reads about two orphaned children living in areas with much lower costs of living. Steve has just enough money to support either his one child in the high-cost area or the two other children in low-cost areas. There is no way he can support all three children.

They were then told to indicate agreement with the following statement:

All things considered, if Joe can either support his one child (in the high-cost area) or two other children (in the low-cost areas), he should support the two children.

For the *intra-agent* version, subjects received the following:

Steve has three young children. One child lives in an area with a high cost of living; his other two children live in areas with much lower costs of living. Steve has just enough money to support either the one child in the high-cost area or the two children in the low-cost areas. There is no way he can support all three children.

They were then told to indicate agreement with the following statement:

All things considered, if Joe can either support his one child (in the high-cost area) or his two other children (in the low-cost areas), he should support the two children.

Again the cases provoked very different responses. The mean response in the inter-agent version is 2.37; the mean response in the intra-agent version is 4.94. The difference between the conditions is again significant (\(t (35) = 5.915, p<.001\)).
instances of promise breaking. Again, some subjects received an *inter-agent* version and the others received an *intra-agent* version. The inter-agent version goes as follows:

Bill promised two sick friends that he would arrange to deliver their homework to them on Saturday. But his original plan fell through and the only alternative is to ask his classmate John to deliver the homework to them on Saturday. But John's parents told him that the only place he can go on Saturday is to the library. If John lies and tells his parents that he is going to the library, he can deliver the homework to Bill's friends. Then Bill will have kept his promises.

Subjects were asked to indicate agreement with the statement, “All things considered, John should lie so that Bill can keep his promises.” As before, subjects tended to disagree with this statement. They thought that John *shouldn’t* lie to enable Bill to keep his promises. The situation was different for the *intra-agent* case, however. Here is that version:

John promised two sick friends that he would deliver their homework to them on Saturday. But John's parents told him that the only place he can go on Saturday is to the library. If John lies and tells his parents that he is going to the library, he can deliver the homework. Then he will have kept his promises.

For this case, subjects were presented with the statement: “All things considered, John should lie so that he can keep his promises.” For this version, subjects tended to agree that John *should* lie so that he can keep his promises.5

Once again, then, we find Nozickian responses for the inter-agent case. People tended to reject the idea that an agent should commit one moral violation so that others do not commit more extensive moral violations. However, the situation is quite different for the intra-agent case. In that case, people tended to say that an agent should commit one moral violation to avoid committing more extensive violations of another sort.

---

5 The mean response in the inter-agent version is 3.57 (out of 7); the mean response in the intra-agent version is 4.70. The difference between the conditions is quite significant (*t* (41) = 2.247, *p*<.05).
§5 Discussion

Nozick uses cases like the magistrate and the mob to motivate the view that (at least certain) moral principles function as side constraints, and not as considerations that enter into cost-benefit calculations. However, cases like the magistrate and the mob, as noted earlier, confound cost-benefit calculation and inter-agency. The cases ask whether an agent should rely on cost-benefit considerations in deciding whether to commit a moral violation to preclude others from committing greater violations. Our experiments show that, once these factors are deconfounded, people are much more willing to engage in cost-benefit analyses. This undercuts Nozick’s brief on behalf of side constraints. In every case we used, participants in the intra-agent condition were much more likely than participants in the inter-agent condition to say that the right thing for the agent to do is violate a moral principle in order to avoid more violations of moral principles. Moreover, the explanations that subjects offered helped to confirm the importance of the confound in Nozick’s cases. In the inter-agent cases, subjects often emphasized the fact that the case involves different agents. One subject wrote, “You shouldn’t break promises. David should have protected his own promise.” Another subject who maintained that John shouldn’t lie to enable another to keep his promises wrote, “It’s not John’s problem.” By contrast, when we turn to the intra-agent cases, we find an emphasis on calculation. For instance, one subject in this condition wrote, “If you can’t fulfill all of your obligations, you may as well fulfill as many as you can.” Once we cast the dilemmas in terms of an intra-agent decision, we find that people do not treat the moral principles as side constraints.

Our findings also suggest that the argument against deontology sketched in section 2 moves too quickly. According to that argument, deontology is defective because it is insensitive to costs and benefits. But our results suggest that while commonsense rebels against utilitarian calculation, commonsense does not rebel against calculation tout court. In fact, people are quite willing to engage in calculation of costs and benefits, so long as the calculation is conducted for a particular agent. In all of our intra-agent cases, subjects judged on the side of permitting one to violate a moral rule to minimize one’s own moral violations, while in all of our inter-agent cases, subjects denied that violating a moral rule to minimize moral violations in general is the right thing to do. Thus, utilitarians can’t argue that deontological prohibitions operate as crude and simplistic psychological alarms. Like Nozick’s case for side constraints, the utilitarian arguments fail to recognize the actual character of our moral sensitivity to costs and benefits; in particular, it fails to recognize that people
do not generally treat moral prohibitions as side constraints that exclude any cost-benefit considerations.\(^6\)

The potential relevance of intra-agent considerations for debates over utilitarianism has not been ignored in the recent literature. Of particular significance is recent work on whether deontological considerations can be ‘consequentialized’ (Dreier 1993; Portmore 2007, 2009). To consequentialize a moral theory, James Dreier writes:

> We merely take the features of an action that the theory considers to be relevant, and build them into the consequences. For example, if a theory says that promises are not to be broken, then we restate this requirement: that a promise has been broken is a bad consequence. Notice that the weighting is not yet specified. If the theory under consideration includes an absolute side constraint against promise-breaking, then we have the consequentialist version give a lexically prior negative weight to promise-breaking (1993, 23).

Dreier thinks that by using this approach, all standard theories can be classified as consequentialist (24). It is important to note that this is a \textit{formal} point, that any standard nonconsequentialist theory can be recast as a consequentialist theory that yields exactly the same recommendations (see, e.g., Portmore 2007, 39-40). There is an ongoing debate over whether the consequentializing approach can provide a catch-all consequentialist standard of right action (see, e.g., Schroeder 2006, 2007; Portmore 2007), but in any case, our experiments have a different target. We are not primarily interested in the formal question--Can deontic theories be recast as consequentialist?; rather, we are interested in a substantive question--What is the \textit{actual} character of our commonsense normative ethics?\(^7\)

---

\(^6\) Although our results suggest that people do not treat all moral principles as side constraints, the results leave open the possibility that \textit{some} moral principles are treated as side constraints. For instance, people might not be willing to allow cost-benefit analyses for acts like rape and torture, regardless of whether the dilemma is cast in inter- or intra-agent terms. People might treat such actions as fully prohibited by side constraints. So, even if some non-utilitarian judgments are driven by alarm bells, this doesn’t cast a pall over the entire sphere of non-utilitarian judgment.

\(^7\) While our findings suggest problems with both the side-constraint view and the standard version of act-utilitarianism, they do not directly speak to whether or not commonsense morality is consequentialist in character.
Our results show that people engage in calculation when what’s at stake are their own moral violations. These findings can be usefully put in terms of the distinction between agent-centered and agent-neutral theories (e.g. Dreier 1993). According to Dreier’s view, an agent-centered theory aims at maximizing the good, though it conceives of the ‘good’ differently from the standard utilitarian approach, which conceives of the good in an agent neutral way. As Dreier draws the distinction, agent-neutral theories give the same advice or aims to everyone, while agent-centered theories give some advice or aims that include indexicals – that is, in some situations the theory may give me advice that conflicts with what the same theory advises you to do (22). The results of our studies suggest that people embrace an agent-centered, rather than agent-neutral, approach. For the studies suggest that when telling a lie would lead to fewer broken promises on the part of Frank, for instance people would give different advice to Frank (lie!) than it will to others (don’t lie!).

The results of our study also bear on the debate over whether deontic judgments are best explained by agent-centered or victim-centered concerns. Frances Kamm (1989) argues that the judgment that it is wrong for someone to kill in order to prevent five others from being killed (inter-agent) is first and foremost grounded in concern for the victim, not the agent. It is wrong for the agent to kill in this case not because the agent should avoid being a killer, but because his potential victim has a constraining right – a right that constrains the agent’s actions. To adjudicate these competing explanations, Kamm appeals to an intra-agent case. She states that it is wrong for the agent to kill the one even if he himself is responsible for the plight of the five others, and will be a killer of the five if he does not now kill the one (1989, 255). According to Kamm, the victim-centered explanation, but not the agent-centered one, can accommodate the judgment that killing is wrong in both types of cases. Indeed, Kamm suggests this is the proper explanation not just for judgments about killing, but for deontic judgments more generally. The results of our study suggest, however, that people’s judgments are more in-line with the agent-centered explanation than with the victim-centered one. Although our cases do not speak to judgments concerning killing, the findings in our intra-agent cases show that subjects judge that agents should act so as to minimize their own moral violations. If Kamm’s victim-centered explanation for commonsense deontic judgments were correct, we would not expect the asymmetry in judgments between inter-agent and intra-agent cases that we do in fact observe.8

8 We thank Doug Portmore and Matt Bedke for drawing our attention to this issue. Portmore (forthcoming) himself has survey data that provides additional evidence against the victim-centered
§6 Agent-centered ethics

We’ve argued that people’s judgments about dilemmas are at odds with both strict utilitarian and strict side-constraint views. Instead, people’s judgments reflect a sensitivity to moral costs and benefits that is most apparent when the moral considerations are agent-centered. What kind of philosophical ethics can make sense of these results? One plausible candidate is Bernard Williams’ view that an essential part of morality is the connection between the agent’s action and her identity—her categorical desires, concerns, and projects. According to Williams (1973), we shouldn’t ask a person to set aside the desires and projects that form a central part of his or her identity. To do so would be to assault the person’s integrity. People naturally care about the shape of their own moral lives. Williams suggests that this is perfectly appropriate, and that a central shortcoming of utilitarianism is its blindness to such agent-centered concerns. Thus, Williams writes:

It is absurd to demand of such a man, when the sums come in from the utility network—which the projects of others have in part determined—that he should just step aside from his own project and decision and acknowledge the decision which utilitarian calculation requires. It is to alienate him in a real sense from his actions and the source of his action in his own convictions. It is to make him into a channel between the input of everyone’s projects, including his own, and an output of optimific decision; but this is to neglect the extent to which his actions and his decisions have to be seen as the actions and decisions which flow from the projects and attitudes with which he is most closely identified (Williams 1973, 116).

Williams illustrates how utilitarians neglect personal integrity with the case of Jim the explorer, who stumbles across a group of 20 natives who are being held hostage. Pedro, the man in charge, tells Jim that he will release 19 of the hostages if Jim kills one of them. If Jim refuses to kill the hostage, then Pedro will kill all 20 hostages himself. Since 20 persons being killed is far worse than one being killed, utilitarianism dictates that Jim should kill the hostage. Even if we agree that killing the one is the right thing for Jim to do, Williams maintains that the utilitarian completely

account. Portmore (forthcoming) also outlines an agent-centered account that aims, pae Kamm, to accommodate deontic verdicts in intra-agent cases.
neglects a morally relevant consideration – in one available action Jim kills the hostage, while in the other available action someone else kills the hostages. What the utilitarian ignores is the idea that “each of us is specially responsible for what he does, rather than for what other people do” (Williams 1973, 99). Our findings in the inter-agent cases dovetail nicely with this point of Williams.

For in the intra-agent cases, people think that agents should behave in accordance with cost-benefit analyses when the moral costs and benefits are within the realm of an agent’s own moral life. One might gloss this as the idea that agents should be concerned with the shape of their own moral lives. And while Williams does not discuss intra-agent cases, his notion of taking special responsibility for one’s own actions certainly seems compatible with engaging in cost-benefit analysis at the intra-agent level in order to minimize one’s own bad actions.

If it’s true that we have a special responsibility for our own actions, and that this drives our judgments about intra-agent cases, this suggests a further experiment. Our previous experiments show that agents express a forward-looking concern for minimizing their own moral violations. A concern for one’s own integrity may also lead an agent to commit a moral violation in order to recompense for a past violation of hers. Do people think that it’s more appropriate to commit a moral violation in order to compensate someone when they bear special responsibility for the loss than when they don’t bear such responsibility? In our final experiment, we presented half of the subjects with a scenario in which an agent can do something morally untoward (breaking a promise) in order to help an accident victim for whom the agent has no special responsibility. The scenario goes as follows:

Mark reads an article in the paper in which an accident victim is interviewed. The victim was involved in a minor car accident that was the fault of the other party. Since there was no damage to the cars and no injuries apparent at the time, the two parties did not exchange information. It turns out the victim sustained severe whiplash that was not apparent at the time of the accident; now he cannot fully afford his costly treatment. The victim says he is

---

9 Rawls (1971) has also raised worries about the lack of agency in utilitarianism. In A Theory of Justice he states, "[The utilitarian] view of social co-operation is the consequence of extending to society the principle of choice for one man, and then, to make this extension work, conflating all persons into one through the imaginative acts of the impartial spectator. Utilitarianism does not take seriously the distinction between persons" (27).
not interested in pursuing legal recourse against the other driver. Mark feels bad for the victim and considers sending him enough money to help him cover his treatment costs. But the only spare money Mark has is money he has been saving to take his family to a concert, which he has long promised to them. If Mark breaks his promise to his family, then he can afford to send a check to the accident victim to cover his medical expenses.

Subjects were then asked to indicate agreement with the following claim:

All things considered, Mark should break his promise to his family so he can send a check to the accident victim.

The other condition, the *special responsibility* condition, is exactly the same except for an additional sentence that makes the agent specially responsible for the accident victim (emphasis not in the experimental materials).

Mark reads an article in the paper in which an accident victim is interviewed. The victim was involved in a minor car accident that was the fault of the other party. Since there was no damage to the cars and no injuries apparent at the time, the two parties did not exchange information. It turns out the victim sustained severe whiplash that was not apparent at the time of the accident; now he cannot fully afford his costly treatment. The victim says he is not interested in pursuing legal recourse against the other driver. As he reads the rest of the article, Mark realizes that he was the one who caused this accident. Mark feels bad for the victim and considers sending him enough money to help him cover his treatment costs. But the only spare money Mark has is money he has been saving to take his family to a concert, which he has long promised to them. If Mark breaks his promise to his family, then he can afford to send a check to the accident victim to cover his medical expenses.

People were significantly more likely to say that Mark should break his promise to help the accident victim when Mark bore responsibility for the accident than when he didn’t.¹⁰

¹⁰ In the special responsibility condition, participants tended to say that Mark should break his promise (average response of 4.47 on a 7 point scale); in the other condition, participants tended to say that Mark shouldn’t break his promise (3.44 on a 7 point scale). The differences was statistically significant (*t*(62) = 2.072, *p* < .05).
Thus our results suggest that at least a part of commonsense ethics conforms to Williams’ ethical vision. Lay judgments suggest that agents should be more concerned with what they do, and with what they have done, than with what others do or have done. People tend to think that what one is morally responsible for is at least partly relative to the agent in question — and this resonates with Williams’ central criticism of utilitarianism in favor of an ethics of integrity. This, of course, fits with the other aspects of the agent-centered ethic outlined thus far. Commonsense morality, our evidence suggests, gives different advice to agents depending on their relation to the outcomes in question. Further, our findings suggest such advice stems from considerations centered on the agent, rather than the victim. The agent-centered ethic sketched here, though far from a complete theory, can systematize the pattern of judgments we’ve observed.

§7 Conclusion

We have suggested that familiar cases used to motivate the side-constraint view of moral rules are inattentive to the ways in which moral considerations operate differently in intra-agent and inter-agent scenarios. Nozick generalizes from the magistrate and the mob case – an inter-agent case – to his side-constraint view concerning moral rules. However, in our study we found that while subjects generally do not subject moral rules to cost-benefit analysis in inter-agent cases, they do perform such calculations in intra-agent cases. This finding undercuts the intuitive support that Nozick recruits for his defense of a broadly side-constraint approach to morality. People are more sensitive to cost-benefit considerations than the side-constraint view suggests. Though people engage in a certain kind of calculation in intra-agent cases, this is not enough to provide intuitive support for a general utilitarian kind of calculation, which demands impartial inter-agent weighing of costs and benefits in all cases. Our experimental findings suggest that commonsense occupies a middle ground between these two views. Moral rules do not act as crude alarm bells, impervious to calculation, nor are they subject to the comprehensive inter-agent style of calculation characteristic of utilitarianism. Rather, people’s deliberations involving moral rules are sensitive to the distinction between inter- and intra-agency. This provides some intuitive support for an agent-centered ethic according to which it is morally appropriate for agents to try to minimize their own moral violations and to take special responsibility for the shape of their own moral lives.
References


