

UC Berkeley

ISUS-X, Tenth Conference of the International Society for Utilitarian Studies

Title

"Utility, Determinism, and Possibility: Context to the Rescue"

Permalink

<https://escholarship.org/uc/item/67c2s2zq>

Author

Norcross, Alastair J.

Publication Date

2008-08-20

Utility, Determinism, and Possibility: Context to the Rescue

Alastair Norcross—University of Colorado at Boulder

Abstract:

Determinism is thought to pose a problem for moral responsibility to the extent that we agree with the principle that someone is only to be held morally responsible for an action if s/he could have done otherwise. The worry, of course, is that if determinism is true, nobody could ever have done otherwise. Utilitarians might seem to be in a better position than other, less enlightened, theorists in this regard. Holding someone responsible, they point out, and related notions such as praise, blame, punishment and reward, are all actions that themselves can be assessed in terms of their consequences. So, the question of whether to hold someone responsible for an action is to be settled by reference to the consequences of the act of holding someone responsible. Whether someone could have done otherwise is, at best, indirectly related to the question of whether and how to hold them responsible. Similarly, the question of whether an act is right or wrong is simply a matter of whether the act was optimal, and has nothing to do with whether the agent could have done otherwise. The problem with this response, of course, is that, if determinism is true, every action is both optimal and pessimal. Every action is both the best and the worst of all the acts that are possible for the agent, because every action is the only action that is possible for the agent. This also applies to the actions of holding responsible, praising, blaming, etc. The solution is to appeal to the conversational context of praising, blaming, judging right and wrong, holding responsible, and the like. Even if, strictly speaking, an agent couldn't have done otherwise, conversational context may select certain counterpossible alternatives as the relevant ones with which to compare the

action. We may, therefore, be able to make sense of a negative (or positive) judgment of an action based on a comparison of the action with an alternative that was not, strictly speaking, available to the agent.

1. The threat of determinism

Determinism is thought to pose a problem for moral responsibility to the extent that we agree with the principle that someone is only to be held morally responsible for an action if s/he could have done otherwise. The worry, of course, is that if determinism is true, nobody could ever have done otherwise. Utilitarians might seem to be in a better position than other, less enlightened, theorists in this regard. The standard utilitarian response to the possibility of determinism (or indeterminism), well articulated by Sidgwick, is twofold. First, holding someone responsible, and related notions such as desert, praise, blame, punishment and reward, are all actions that themselves can be assessed in terms of their consequences.

..the Determinist can give to the terms "ill-desert" and "responsibility" a signification which is not only clear and definite, but, from an utilitarian point of view, the only suitable meaning. In this view, if I affirm that A is responsible for a harmful act, I mean that it is right to punish him for it; primarily, in order that the fear of punishment may prevent him and others from committing similar acts in future. (ME Bk. I ch. V sec. 4)

So, the question of whether to hold someone responsible for an action is to be settled by reference to the consequences of the act of holding someone responsible. Whether someone could have done otherwise is, at best, indirectly related to the question of

whether and how to hold them responsible. But what of the question of what makes an act right or wrong? If we say that someone is morally responsible for an act just in case it is right to punish him for it, we need to know what makes such an act of punishing right, as well as what makes *any* act right. The standard utilitarian account, again at least since Sidgwick, is that the question of whether an act is right or wrong is simply a matter of whether the act was optimal. An act is right just in case it is the best act (in terms of its consequences) that the agent could have performed. Of course, this seems to bring in the question of whether the agent could have done otherwise. If determinism is true, it might be thought that every action is both optimal and pessimal. Every action is both the best and the worst of all the acts that the agent could have performed, because every action is the *only* action that the agent could have performed. If optimality is a sufficient condition for rightness, all our actions will be right, which entails that all acts of blaming, punishing, holding responsible, etc. will also be right. This would lead to the rather counterintuitive conclusion that every act that is punished is both blameworthy and right!

Sidgwick's response to this is:

As regards action generally, the Determinist allows that a man is only morally bound to do what is "in his power"; but he explains "in his power" to mean that the result in question will be produced if the man choose to produce it. And this is, I think, the sense in which the proposition "what I ought to do I can do" is commonly accepted: it means "can do if I choose", not "can choose to do". (ME Bk. I ch. V sec. 3)

Likewise, G. E. Moore has a similar account of these notions:

all along we have been using the words "can," "could," and "possible" *in a special sense*. It was explained in Chapter I (§§ 17-18), that we proposed, purely for the sake

of brevity, to say that an agent *could* have done a given action, which he didn't do, wherever it is true that he could have done it, *if* he had chosen; and similarly by what he *can* do, or what is *possible*, we have always meant merely what is possible, *if* he chooses. Our theory, therefore, has not been maintaining, after all, that right and wrong depend upon what the agent absolutely *can* do, but only on what he can do, *if* he chooses. (*Ethics*, ch. 6, sec. 3)

So, an act A is right just in case there is no better act B, such that if the agent chose to do B, he would have done B. This allows the consequentialist to talk of the range of actions that are available to an agent in a given situation, even if determinism is true. Actions may be non-trivially optimific, or truly sub-optimific. The standard maximizing conception of rightness can be used to judge actions, as can a satisficing conception, or a scalar conception. The range of available alternatives to a given action are all those actions of which it is true that, if the agent had chosen to perform them, she would have succeeded in performing them. Whether she could have chosen to perform them is irrelevant.

But things are not so simple. Consider a standard Frankfurt-style example designed to undercut the Principle of Alternate Possibilities. George is considering whether to sign into law a statute requiring homosexuals to be branded with a pink triangle on their foreheads and atheists to be branded with a scarlet 'A'. Karl has a completely reliable mind-control device with which he can guarantee that George will choose to sign the law. If the device detects that George is about to choose to veto the law, it will make him choose to sign it. Otherwise, the device is inactive. In the standard version of this example, George chooses without any help from Karl's device. That is,

George chooses to sign the bill, and he would have so chosen, even if Karl's device hadn't existed. The example is supposed to pump our intuitions that George is morally responsible for signing the bill (or choosing to sign it), even though he couldn't have done otherwise. Likewise it could pump our intuitions that George's act of signing is wrong.

Now consider the case in which the device actually operates to ensure that George chooses to sign the bill. That is, the device detects that George is about to choose to veto the bill, so it intervenes to make George choose to sign it. Now, although George chooses to sign the bill, it is no longer true that he would have so chosen even if Karl's device hadn't existed. This version certainly seems to pump our intuitions that George is *not* morally responsible in this case. Leaving aside intuitions about moral responsibility, what should we say about how George's behavior (in either version) compares with his possible alternatives? Do we say that it is suboptimal, because, if he had chosen to veto the bill, homosexuals and atheists wouldn't have been persecuted, at least not quite so much?

But how do we evaluate that counterfactual? The closest world in which George chooses to veto the bill is one in which Karl's device doesn't work. Given that Karl's device is completely reliable, the closest world in which it doesn't work may well be one in which it doesn't even exist. Given that Karl's device has been highly instrumental in gaining political power for George, if the device hadn't existed, or if it had been unreliable, George wouldn't have had political power, and so wouldn't have been in a position to sign or veto the law in question. In fact, given that Karl's device has been highly instrumental in gaining political power for those who wrote and passed the law in

the first place, if the device hadn't existed or had been unreliable, there would have been no such law for *anyone* to sign or veto. So, the closest world in which George chooses to veto the law may be very far indeed from the actual world in which George chooses to sign the law. Given the ways in which the world would have had to have been different in order for George and all the homophobic theocrats to have gained power without Karl's device, it may even be the case that that world with the law vetoed is worse than the actual world with the law signed. (The level of homophobic and religious persecution in the distant veto world may be so high, even without the branding law, that not only the overall wellbeing, but even the wellbeing of homosexuals and atheists may be higher in the actual signing world.)

If the closest world in which George chooses to veto the law is as far from the actual world as all that, why would we claim that vetoing is a relevant alternative with which to compare George's act of signing? Compare this case with a straightforward case of physical disability limiting a range of choices. Suppose that Mary is a kindergarten teacher. An explosion in the kindergarten causes a joist to sever her left leg below the knee, and traps two children, Bill and Ben, in a burning room. Mary quickly wraps a tourniquet around her leg, and hops into the room to save Bill, who is closer to the door. However, by the time she has got Bill to safety, Ben is dead. If she had had both legs, she could have run into the room and had time to save both Bill and Ben. However, given her recent loss of her leg, her saving Bill is the best she could do, and is pretty heroic to boot. It simply wasn't in her power to save both, because it wasn't in her power to run (as opposed to hop) into the room. But what about the counterfactual "if she chose to run into the room, she would have succeeded in running into the room"?

That may well be true. After all, Mary wouldn't have chosen to run into the room, unless she hadn't lost a leg. She is level-headed enough not to choose to do something that she knows full well she cannot do. The world in which Mary doesn't lose a leg, perhaps because she moved just before the joist fell, and does choose to run into the room to save both Bill and Ben, is much closer to the actual, hopping world than the closest George veto-choosing world is to the actual George signing-choosing world. But the fact that there is a relatively close world in which Mary doesn't lose her leg and thus does choose to run doesn't ground the claim that it was in Mary's power to run in the actual world.

It might be objected at this point that I am unfairly bringing in details of what the world would have had to have been like in order for George to have vetoed the law. Sidgwick's suggestion is about the meaning of "in his power", and is intended to apply to the question of what actions are in our power, even if determinism is true. We should simply stipulate that we ban backtracking counterfactuals, and hold everything about the actual the world constant, except for the choice itself. In this case, despite Karl's device both existing and operating, George chooses to veto the law. Quite apart from the considerable whiff of adhocery about this move, it leaves open the question of how to assess the option of vetoing the law. Clearly, we are not supposed to hold *everything* about the actual world except for the choice the same. We are supposed to alter things *after* the choice. So, as a result of choosing to veto the law, George actually does veto the law. As a result of George vetoing the law, it doesn't go into effect, and many homosexuals and atheists are persecuted less than in the actual world. This is what is supposed to ground our judgment that George's act of signing the law is suboptimal. But what of Karl's reaction to what he sees as a violation of natural law? Since we can't

suppose that Karl's device malfunctions in any way prior to George choosing to veto the law, Karl may well conclude from the fact that George chooses to veto that the natural order has broken down. Perhaps he takes this as a sign of the end times. The rapture has occurred, and he is still down on earth! Who knows what such beliefs would push him to do? We certainly can't be sure that the results would be better overall than those of George signing the original law. Furthermore, in the version of the George example in which Karl's device detects that George is about to choose to veto, and so works its neurophysiological magic to ensure that George chooses to sign, are we really prepared to say that it was in George's power to sign?

A more humdrum example can also illustrate a problem with Sidgwick's suggestion. Frances Howard-Snyder considers an example in which she is playing chess against Karpov. Suppose that something really important depends on whether she beats Karpov. Perhaps some innocent lives will be saved just in case she beats Karpov, and she knows it. Given that, it would clearly be wrong of her not to beat Karpov, if she had it in her power to beat him. However, we are all supposed to agree, Frances cannot beat Karpov. He is the world chess champion, and she is a lowly philosophy professor from Bellingham. No matter how hard she tries, she cannot beat him. But this is not because he is unbeatable. After all, Big Blue has beaten him, as have some very good human chess players. There are some sequences of moves, such that those sequences would result in beating Karpov. Each of those sequences consists of moves that Frances could make, if she chose to do so. Call one such sequence 'A'. If Frances executes A, she beats Karpov. Furthermore, if she chooses to execute A, she can execute A. So, why isn't it in her power to beat Karpov?

I suggest that the solution to these problems for the consequentialist is to appeal to the conversational context of praising, blaming, judging right and wrong, holding responsible, and the like. Even if, strictly speaking, an agent couldn't have done otherwise, conversational context may select certain counterpossible alternatives as the relevant ones with which to compare the action. We may, therefore, be able to make sense of a negative (or positive) judgment of an action based on a comparison of the action with an alternative that was not, strictly speaking, available to the agent. In what follows, I will first sketch an independent motivation for a consequentialist to embrace contextualist accounts of various ethical terms. Next, I will briefly explain the contextualist approaches to these terms. Finally, I will explain how this form of contextualism can be applied to the problems of free will and responsibility.

2. The motivation for contextualism

I have argued elsewhere¹ that consequentialism is not fundamentally concerned with such staples of moral theory as rightness, duty, permissibility, obligation, moral requirements, goodness (as applied to actions), and harm. In fact, I have argued that the standard consequentialist accounts of these notions are either indeterminate (in the case of the latter two) or redundant. What is fundamental to a consequentialist ethical theory is a value theory, for example hedonism or some other form of welfarism, and the claim that

¹ "Good and Bad Actions", *The Philosophical Review*, Vol 106, No. 1; January 1997, pp. 1-34; "A Consequentialist Case for Rejecting the Right", *The Journal of Philosophical Research*, Vol. 18; 1993, pp. 109-125, co-authored with Frances Howard-Snyder; "Reasons and Demands: Rethinking Rightness", in James Dreier (ed.) *Blackwell Contemporary Debates in Moral Theory*, 2006; "Harming in Context", *Philosophical Studies*, Vol 123, Nos 1-2, March 2005; "Scalar Act-Utilitarianism", in Henry R. West (ed.) *Blackwell Guide to Mill's Utilitarianism*, 2006.

the objects of moral evaluation, such as actions, characters, institutions, etc. are compared with possible alternatives in terms of their comparative contribution to the good. For example, one action is better than another, just in case, and to the extent that, the world that contains it is better than the world that contains the other from the time of the choice onwards. Furthermore, our (moral) reasons for choosing between alternative actions, institutions, etc. are essentially comparative, and correspond to the comparative consequential value of the options. I might have a better reason for choosing to do A than to do B, and better by a certain amount, but neither reason is either good or bad *simpliciter*. So, if all a consequentialist moral theory supports at the fundamental level are comparative evaluations of actions, characters, institutions (and thus also comparative reasons for choosing among them), what, if anything, does it have to say about such notions as right and wrong, duty, obligation, good and bad actions or harm? There seem to be two main options, one of which is the form of ethical contextualism which is the main focus of this paper. The other, which I will only briefly mention here, is a form of eliminativism, combined with an error theory regarding our common usage of these terms. The consequentialist could simply say that there's no such thing as right and wrong actions, good and bad actions, harmful actions, etc. It doesn't, of course, follow from this that "anything goes", if that is taken to mean that everything is permissible, and so, for example, it's perfectly permissible to torture innocent children. Just as no actions are either right or wrong, none are permissible or impermissible either. Neither does it follow that anything goes, if that is taken to mean that morality has nothing to say about actions. The action of torturing an innocent child will almost certainly be much worse than many easily available alternatives, and thus strongly opposed by moral reasons when

compared with other options. It does, however, follow that descriptions of actions (or characters, or institutions) as being right or wrong, good or bad, harmful, required, permissible, and the like are all mistaken (either false or meaningless). This might seem to be a rather uncomfortable result. We can understand how some, perhaps many, claims about the rightness or goodness or permissibility of actions are mistaken, but *all claims*? Is it plausible that we have *all* been mistaken all this time? I don't find this possibility particularly implausible. Similar things may well be true for certain areas of theological or scientific discourse. If there is no god, for example, all claims about what god loves or hates are mistaken (either false or meaningless). Similarly, much scientific discourse assumes the existence of entities that may turn out not to exist. It might, perhaps, be argued that the situation is different for morality. While theology and fundamental physics is unashamedly concerned with unobservable, or at least difficult to observe, entities, morality is concerned with everyday properties that require little or no expertise to discern. Although I don't find such considerations particularly compelling, I do want to explore how a consequentialist can accommodate some of the commonly-accepted moral properties, despite excluding them from the fundamental level of the theory.

What I propose is a form of contextualist analysis of the relevant moral terms, similar in form to some recent contextualist approaches to the epistemological notions of knowledge and justification. Roughly, to say that an action is right, obligatory, morally required, etc. is to say that it is at least as good as the appropriate alternative (which may be the action itself). Similarly, to say that an action is good is to say that it resulted in a better world than would have resulted had the appropriate alternative been performed. To say that an action harmed someone is to say that the action resulted in that person being

worse off than they would have been had the appropriate alternative been performed. In each case, the context in which the judgment is made determines the appropriate ideal or alternative. I will illustrate first with the cases of good actions and harmful actions, and then say a little about a contextualist analysis of ‘right’.

3. Problems with Non-contextualist Accounts of Good and Harm

In order to explain (and motivate) a contextualist account of good actions and harmful actions, I will briefly explain why satisfactory noncontextualist accounts of such notions are not available to the consequentialist². If the goodness of an action is to be a consequentialist property, something like the following account suggests itself:

G An act is good iff it produces more goodness than badness; an act is bad iff it produces more badness than goodness.

The general idea expressed in **G** is used by philosophers, both consequentialist and non-consequentialist³, though not necessarily as an explicit account of good and bad actions. But what does it mean to produce more goodness than badness, or, to put it another way, to have consequences that are on balance good?

The obvious answer is that for an action to have on balance good consequences is for it to make a positive difference in the world, that is, to make the world better. But better than what? A first attempt is to say better than it was before the action. But this clearly won’t do. To see this, consider an example in which the world contains only two

² For fully detailed accounts, see “Good and Bad Actions”, and “Harming in Context”, Op. Cit.

³ For specific examples, see “Good and Bad Actions” op. Cit., 5-7.

sentient beings, Agent and Patient. Patient is terminally ill. His condition is declining, and his suffering is increasing. Agent cannot delay Patient's death. The only thing she can do is to slow the rate of increase of Patient's suffering by administering various drugs. The best available drugs completely remove the pain that Patient would have suffered as a result of his illness. However, they also produce, as a side-effect, a level of suffering that is dramatically lower than he would have experienced without them, but significantly higher than he is now experiencing. So the result of administering the drugs is that Patient's suffering continues to increase, but at a slower rate than he would have experienced without them. The very best thing she can do has the consequence that Patient's suffering increases. The world is worse after Agent's action than it was before, but Agent's action is clearly not on that account bad. In fact, inasmuch as a consequentialist is inclined to make a judgment about the action's goodness, she would say that it is good.

In evaluating actions, a consequentialist compares states of affairs, not across times, but across worlds. The reason why it seems as if Agent's action is good is that it does make the world better, not better than it was, but better than it *would have been* if the action hadn't been performed. This suggests the following account of good actions:

GC: An act A is good iff the world would have been worse if A hadn't been performed; A is bad iff the world would have been better if A hadn't been performed.

This explains why Agent's action is good. If she hadn't administered those drugs to Patient, Patient would have suffered even more. But this is an easy case, which hides a crucial problem with GC. According to GC, whether an action is good or bad depends on

what the world would have been like if it hadn't been performed. So, what would the world have been like, if Agent had administered those drugs to Patient? That depends on what Agent would have done instead. She might have tried a different course of treatment, which was less effective. She might have simply sat and watched while Patient's suffering increased. She might have tried a different course of treatment that actually increased the rate of increase of Patient's suffering (either intentionally or not). In this case, we don't need to know precisely what Agent would have done instead, because we know that she did the best she could, and thus that the world would have been worse, if she had done *anything* else.

But other examples are not so easy. Consider the following:

Button Pusher. Agent can push any one of ten buttons (labeled '0' through '9'), killing between none and nine people, or push no button at all, with the result that ten people die. No button is any more difficult to push than any other, nor is there any pressure (either physical or psychological) exerted on Agent to push any particular button.

Suppose that Agent pushes the button labeled '9', with the result that nine people die. Intuitively, this seems like a pretty bad action. However, suppose also that Agent is highly misanthropic, and wants as many people as possible to die. Her initial inclination was to press no button at all, so that all ten would die. She also enjoys being personally involved in the misfortunes of others, however, and believes that pressing a button would involve killing, whereas refraining from pressing any button would involve 'merely' letting die, which, from her misguided perspective, is less personally involving. She struggled long and hard over her decision, weighing the advantage of one more death

against the disadvantage of less personal involvement. She never contemplated pressing any button other than '9'. It's clear, then, that if Agent hadn't pressed '9', she would have pressed no button at all. So the world would have been worse, if she hadn't pressed '9'. But this doesn't incline us to judge her action to be good.

Although *Button Pusher* might suggest that anything less than the best action is bad, we are not likely to endorse that as a general principle. Consider:

Burning Building. There are ten people trapped in a burning building. Agent can rescue them one at a time. Each trip into the building to rescue one person involves a considerable amount of effort, risk and unpleasantness. It is possible, albeit difficult and risky, for Agent to rescue all ten.

Suppose that Agent rescues nine people, and then stops, exhausted and burned. She could have rescued the tenth, so doesn't do the very best she can, but do we really want to say that her rescue of nine people wasn't good (was actually bad)?

None of the different interpretations of **GC** can provide the consequentialist with a satisfactory account of what it is for an action to be good. The intuition on which they are based is that a good action makes the world better. The difficulty lies in producing a general formula to identify the particular possible world (or worlds), than which the actual world is better, as a result of a good action. Any unified theory requires a way of fixing the contrast point, but the contrast point varies from situation to situation. Part of the problem is that our intuitions about the goodness or badness of particular actions are often influenced by features of the context that it would be difficult to incorporate into a general account.

Consider now the consequentialist approach to harm:

HARM An act A harms a person P iff P is worse off, as a consequence of A, than she would have been if A hadn't been performed. An act A benefits a person P iff P is better off, as a consequence of A, than she would have been if A hadn't been performed.⁴

It is easy to see that the same problems that apply to the consequentialist account of good and bad actions apply to the consequentialist account of harmful and beneficial actions. The following example will illustrate: suppose you witness the following scene at Texas Tech University: A member of the Philosophy department, passing Bobby Knight on campus, waves cheerily and says “Hey, Knight.” Bobby Knight, turning as red as his sweater, seizes the hapless philosopher around the neck and chokes her violently, while screaming obscenities. By the time Bobby Knight has been dragged away, the philosopher has suffered a partially crushed windpipe and sustained permanent damage to her voicebox, as a result of which she will forever sound like Harvey Fierstein.

Has Bobby Knight’s act harmed the philosopher? The intuitive answer is obvious, and HARM seems to agree. The philosopher is much worse off than she would have been had Bobby Knight not choked her (unless, perhaps, she has always wanted to sound like Harvey Fierstein). But suppose we discover that Bobby Knight has recently been attending anger management classes. Furthermore, they have been highly successful in getting him to control his behavior. When he becomes enraged, he holds himself relatively in check. On this particular occasion (only the third violent outburst of the day), he tried, successfully, to tone down his behavior. In fact, if he hadn’t been

⁴ Something like this principle is assumed in most consequentialist writing about harmful and beneficial actions. See, for example, Derek Parfit, *Reasons and Persons*.

applying his anger management techniques, he wouldn't have choked the philosopher, but would rather have torn both her arms from her body and beaten her over the head with them. Since it took great effort on Bobby Knight's part to restrain himself as much as he did, it seems that the closest possible world in which he doesn't choke the philosopher is one in which she is even worse off. HARM, in this case, seems to give us the highly counterintuitive result that, not only does Bobby Knight's act of choking not harm the philosopher, it actually benefits her. HARM also seems to give the result in *Button Pusher* that Agent doesn't harm any of the nine people who die as a result of pushing '9'. They are no worse off than they would have been if she hadn't pressed '9'. If she hadn't pressed '9', she wouldn't have pressed any button, and all ten people would have died.

As with good and bad actions, the consequentialist account of harmful and beneficial actions includes a comparison with an alternative possible world. To harm someone is to make her worse off than she *would have been*. The alternative with which we are to compare the actual action, though, is not always plausibly identified by the counterfactual. Features of the conversational context in which a particular action is being assessed can affect which alternative is the appropriate one.⁵

4. Contextualism about Good, Harm and Right

⁵ Counterfactuals themselves are, of course, infected by context, but not always in the same way as judgments about harm. For example, simply entertaining a counterfactual may change the context in a way in which considering a judgment of harm without explicitly entertaining the counterfactual may not.

Consider first a contextualist account of good action:

G-con An action is good iff it is better than the appropriate alternative.

As examples for which the conversational context is unlikely to change the appropriate alternative, consider again *Button Pusher* and *Burning Building*. Suppose that Agent pushes ‘5’ in *Button Pusher*. It is hard to imagine a conversational context in which anything other than pushing ‘0’ is selected as the appropriate alternative. Pushing ‘5’ would clearly be judged a bad action in just about any plausible conversational context. Now suppose that agent rescues three people in *Burning Building*. In most conversational contexts the appropriate alternative will be rescuing none (or perhaps one), and so the rescue of three will be judged to be good.

Now consider an example for which conversational context might change the appropriate alternative.

Perot. Ross Perot gives \$1000 to help the homeless in Dallas and I give \$100.

In most conversational contexts both of our actions will be judged to be good, because the appropriate alternatives will be ones in which we give no money. But consider again Perot’s donation. Let’s add a couple of details to the case: (i) Perot has a firm policy of donating up to, but no more than, \$1000 per month to charity. (Some months he gives less than \$1000, even as little as nothing at all, but he never gives more than \$1000.) (ii) He had been intending to give \$1000 this month to complete construction on a dam to provide water for a drought-stricken village in Somalia. As a result of Perot’s switching the money this month to the homeless in Dallas, the dam takes another month to complete, during which time twenty children die of dehydration. Now it is not nearly so clear that we should say that Perot’s action was good. A change in the description of the

action might change the appropriate comparison. The extra details about the dam in Somalia make it unclear how to evaluate the action. It is still true that giving the \$1000 to the homeless is better than leaving it in the bank, but it is unclear whether this continues to ground the judgment that Perot's action is good. In fact, it is very tempting to say that Perot did a bad thing by diverting the money from the dam to the homeless. The point here is not just that *learning* the details of the dam in Somalia changes the appropriate comparison. The point is rather that what comparisons are appropriate can change with a change in the linguistic context, even if there is no epistemic change. For example, different descriptions of the same action can make different comparisons appropriate. If we ask whether Perot's *diversion* of the \$1000 from the starving Somalians to the Dallas homeless was good, we will probably compare the results of the actual donation with the alternative donation to the Somalians. If, however, we ask whether Perot's *donation* to the Dallas homeless was good, we may simply compare the donation to the alternative in which the money sits in the bank, even if we know that Perot had previously intended to send the money to Somalia. Perhaps we'll say that the action was good, but not as good as the alternative of aiding the Somalians.

It might be objected at this point that there are theories of action individuation, according to which Perot's *diversion* of the \$1000 from the starving Somalians to the Dallas homeless is not the same action as Perot's *donation* to the Dallas homeless. According to such theories, my example involves a switch from one action to another (spatiotemporally coextensive) one, rather than a mere switch in the way of describing a single action. However, there can clearly be changes in linguistic context that affect the appropriateness of comparisons, without affecting which action is being referred to, on

any plausible theory of action individuation. There may be a change in the appropriate comparison even without a change of action description. Suppose that, just before asking whether Perot's donation to the Dallas homeless was good, we have been discussing his prior intention to give the money to the Somalians. In this context, we are quite likely to compare the actual donation with the better alternative. On the other hand, suppose that, just before asking whether his donation was good, we have been discussing the fact that Perot has made no charitable contributions at all in four of the last six months, and small ones in the other two. In this context, we will probably compare the actual donation with a worse alternative.

Now let's consider a contextualist account of harm:

H-con An action A harms a person P iff it results in P being worse off than s/he would have been had the appropriate alternative been performed.

Many straightforward examples involve actions for which the conversational context is most unlikely to change the appropriate comparison, or at least unlikely to change it so as to produce a different judgment. For example, chancing to encounter you at a philosophy conference, I kill and eat you. It is hard to imagine a conversational context in which the appropriate alternative action is worse for you than being killed and eaten. Likewise, to use a real example, if I say that Booth's shot harmed Lincoln, the context selects, as an appropriate alternative act of Booth, pretty much anything else except shooting Lincoln. It may be true that Booth could have shot Lincoln in such a way as to lead to a much more agonizing death than the one he in fact suffered. This alternative, however, is normally not salient (and may never be). However, it's also a fairly straightforward matter to produce an example for which the appropriate alternative does change with the

conversational context. Sometimes different, equally normal, contexts can render one act a harming or a benefiting. For example, my father writes a will, in which I receive half his estate. This is the first will he has written. Had he died intestate, I would have received all of his estate. Two among his many other options were to leave me none of his estate or all of it. Does my father's act of will-writing harm me or benefit me?

Imagine a conversation focused on my previous plans to invest the whole estate, based on my expectation that I would receive the whole estate. It might be natural in such a context to describe my father's act as harming me. I end up worse off than if he had left me all his estate, which I had expected him to do, either by not making a will at all, or by making one in which he left me the whole shebang. Imagine, though, a different, but equally natural, conversation focusing on my lack of filial piety and the fact that I clearly deserve none of the estate. In this context it may be natural to describe my father's act as benefiting me. After all, he *should* have left me nothing, such a sorry excuse for a human being I was.

At this point an objection may arise. Introducing the previous example, I said that different contexts can render one act a harming or a benefiting. Given that I am talking about harm *all things considered*, how can I claim that one act can correctly be described as both a harming and a benefiting? Wouldn't this be contradictory? Likewise, in discussing Perot's donation to the Dallas homeless, I said that the context in which it is discussed can determine whether the appropriate comparison is with a better or a worse alternative, and thus whether Perot's action is correctly described as good or bad. Again, it seems that I am claiming that one action can be correctly described as both good and bad. Isn't this contradictory? No. In order to see why not, we need to be precise about

what I am committed to. I say that one act can be correctly described *in one conversational context* as good, and can be correctly described *in a different conversational context* as bad. The reason why no contradiction is involved is that a claim of the form ‘act A was good’ can express different propositions in different contexts. (The same point, of course, applies to the contextualist accounts of harm and other moral notions.) On my suggested account of good actions, to claim that act A was good is to claim that A resulted in a better world than would have resulted if *the appropriate alternative* to A had been performed. Given the context-relativity of *the appropriate alternative*, claims about good and bad actions have an indexical element. Just as ‘today is a good day to die’ can express different propositions in different contexts of utterance, so can ‘Perot’s donation to the Dallas homeless was good’.

At this point I should clarify the role of salience in my contextualist account of moral terms. I mean by salience, roughly, the degree to which the participants in a conversational context consciously focus on an alternative. There may be more sophisticated accounts of salience, but this is certainly a common one. Salience often plays a role in determining which alternative the context selects as the appropriate one, but salience may not be the only determining factor. To see this, consider an example that might be thought to pose a problem for my account, if salience is solely responsible for selecting the appropriate alternative.⁶ Imagine a group of comic-book enthusiasts talking about how great it would be if their leader, Ben, had the abilities of Spiderman. After an hour or three of satisfying fantasizing, they are joined by Ben himself, who

⁶ I owe at least the general idea of this example, though not the details, to Ben Bradley. He suggested something like this in discussion as a problem for my account.

apologizes for being late. He explains that he was on his way when his grandmother called him on his cellphone. She had fallen, and she couldn't get up without his help. It took him more than an hour to get to her, because of traffic congestion, during which time she had been lying uncomfortably on the floor. Once he helped her up, though, she was fine. He is sorry that he is late, but the rest of the group, who are also devoted grandsons, must agree that benefiting his grandmother is a good excuse. "Au contraire", reply his friends, that is the "worst excuse ever". He didn't benefit his grandmother at all, but rather harmed her, since he would have reached her a lot sooner, and prevented much suffering, if he had simply used his super spider powers to swing from building to building, instead of inching his way in traffic. Furthermore, he would have reached the meeting on time. Clearly, something is amiss here. Even though the alternative in which Ben swings through the air on spidery filaments is, in *some* sense, salient, it is not thereby the appropriate alternative with which to compare his actual behavior. We can't make an alternative appropriate simply by talking about it, although we may be able to make it salient that way. Perhaps we should add to salience, among other things, a commitment to something like 'ought implies can'. Since Ben cannot swing through the air on spidery filaments, this is ruled out as an appropriate alternative.⁷ I don't here have the time (or the inclination) to give a detailed account of how conversational context determines the appropriate alternative. I suspect that the correct account will be similar to the approach of contextualists in epistemology, such as David Lewis, Mark Heller, and Keith deRose.

Finally, consider a contextualist analysis of 'right':

R-con An action is right iff it is at least as good as the appropriate alternative.

⁷ I owe this suggestion to Julia Driver.

The idea here is that the concept of right action (and duty, permissibility, obligation, and the like) invokes a standard, against which the action in question is judged. The standard maximizing consequentialist theory is a non-contextualist theory of the right, which fixes the standard as optimizing. For the maximizer, the appropriate ideal is always the optimal option. However, the contextualist approach I am suggesting allows the conversational context to affect the standard. It seems likely that most (ordinary) contexts will be sensitive to such factors as difficulty (both physical and psychological), risk, and self-sacrifice in establishing the appropriate ideal. For example, most, if not all, contexts will establish the act of pushing ‘0’ as the appropriate ideal in *Button Pusher*, so that any other action will be judged wrong. *Burning Building*, is a little trickier, but it is hard to imagine many ordinary contexts that set the rescue of *everyone* as the appropriate alternative. The standard criticism of maximizing consequentialism that it fails to accommodate supererogation is based in the intuition that there are cases in which duty, or right action, doesn’t demand maximizing. *Burning Building* seems to be a good example of one. In order to get a context that would set the rescue of all ten people as the appropriate ideal, we could imagine a conversation among committed maximizing consequentialists, or perhaps among proponents of a Christ-as-ideal moral theory, or perhaps it will be enough to imagine a conversation in a philosophy class that has just been presented with maximizing consequentialism. Just as the epistemological contextualist presents classroom contexts as setting particularly demanding epistemic standards, and thus as being ones in which “I don’t know that I have hands” can be uttered truly, so the ethical contextualist can present classroom contexts in which the maximizing alternative determines the truth value of claims of rightness. Of course,

classroom contexts might also set very low standards. A discussion of the demandingness objection to consequentialism might set a pretty lax standard.

It is important to stress that the contextualism I am suggesting is a *linguistic* thesis. I am not suggesting that the rightness (or goodness, etc.) of a particular action can vary with the context in which it is discussed. I am suggesting that a sentence such as “Michael Moore was morally right to describe Bush as a ‘deserter’” may express different propositions when uttered in different contexts. The rightness of Moore’s act (of describing) doesn’t vary with the context in which it is discussed. That is because the context in which the previous sentence was uttered (or read) determined the property picked out by ‘rightness’ *in that context*. Assume that Moore’s act possessed *that* property. If so, no change in linguistic context can change the fact that Moore’s act possessed *that* property. A change in linguistic context can make it the case that a different utterance of ‘rightness’ will pick out a different property.

A contextualist approach to all these notions makes room for them in ordinary moral discourse, but it also illustrates why there is no room for them at the level of fundamental moral theory. If the truth value of a judgment that an action is right or good varies according to the context in which it is made, then rightness or goodness can no more be properties of actions themselves than thisness or hereeness can be properties of things or locations themselves. To be more accurate, since ‘right’ (and the other terms I have discussed) can be used to pick out different properties when used in different contexts, many actions will possess *a* property that can be legitimately picked out by ‘right’ (or ‘good’, ‘harmful’, etc.) and lack many other such properties.

5. Back to determinism and responsibility

So, how does this contextualist approach apply to the examples I discussed earlier, or to the possibility that determinism is true? Consider first Frances's chess match against Karpov. Even though there is at least one sequence of moves, A, that would have beaten Karpov, and Frances could have executed A if she had chosen to do so, given that she didn't know what A was, and couldn't reasonably have been expected to know, no conversational context would select A as an appropriate alternative to whatever sequence of moves Frances did make. That is because our judgments about rightness are constrained by our judgments about abilities, and our judgments about abilities are constrained in at least some ways by our judgments about knowledge. The examples of George signing the bill are a bit more complex. Some contexts may be sensitive to the presence of Karl's device in such a way that it renders George's act not wrong, though it may not be right either. I suspect that most conversational contexts presuppose that such devices aren't present. Consider what would happen if the parties to a conversation concerning an acknowledged wrong action discovered that the agent's choice had been ensured by such a device. Of course, since such devices don't exist, we don't have any hard data on how such a discovery would affect our judgments, but we can make intelligent guesses. Furthermore, there are realistic cases in which we discover that an agent's choice was strongly influenced by mental abnormalities, such as the belief that god told her to do it, that she is actually Napoleon, or that tax cuts for the wealthy will "trickle down" and eventually benefit the poor. Such discoveries often do affect our judgments of the rightness or wrongness of actions, or even of the appropriateness of making such judgments at all. Similarly, it may be a presupposition of all normal

conversational contexts that our actions are not determined. Sidgwick may seem to be suggesting something like this when he says:

Certainly when I have a distinct consciousness of choosing between alternatives of conduct, one of which I conceive as right or reasonable, I find it impossible not to think that I can now choose to do what I so conceive,.. I can suppose that my conviction of free choice *may* be illusory: that if I knew my own nature, I *might* see it to be predetermined that, being so constituted and in such circumstances, I should act on the occasion in question contrary to my rational judgment. But I cannot conceive myself seeing this, without at the same time conceiving my whole conception of what I now call ``my" action fundamentally altered: I cannot conceive that if I contemplated the actions of my organism in this light I should refer them to my "self"---*i.e.* to the mind so contemplating---in the sense in which I now refer them. (ME Bk. I ch. V sec. 3)

Just as a classroom context in which evil demon scenarios are discussed may make denials of even quite simple knowledge, such as “I know that I have hands” true, so may classroom contexts in which the possibility that all our actions are determined may make the denial of moral judgments about our actions true. Well, they may make such denials true, if determinism is true.