CAMBRIDGE
UNIVERSITY PRESS

ARTICLES

Harm, Context, Blame, and Significance: A Response to Eggleston, Sinnott-Armstrong, Mason, and Kagan

Alastair Norcross

Department of Philosophy, University of Colorado Boulder, Boulder, CO, USA
Email: Alastair.Norcross@Colorado.edu

Abstract

Eggleston claims that my account of harm suffers from more problems than his preferred account. I clarify my account, and explain how his account suffers from some of the supposed problems he charges my account with. Sinnott-Armstrong suggests that his contrastivist approach is preferable to my contextualism. I clarify the role of linguistic context, and suggest that our positions are quite close to each other. Mason worries that my scalar approach does not properly accommodate the notions of blame and moral responsibility. I maintain that such notions have only a derivative status, but are nonetheless important, and I suggest fruitful avenues for the scalar consequentialist to pursue. Kagan claims that the addition of a contextualist account of "right" renders my view not importantly different from maximizing or satisficing views. I explain why this is mistaken, and why neither maximizing nor satisficing versions of rightness can explain its supposed moral significance.

Keywords: ethics; harm; blame; responsibility; scalar

Eggleston

Ben challenges my treatment of the notion of harm (as applied to actions) on two major grounds. First, he raises doubts about my proposed contextualist analysis of statements of the form "Act A harmed person P." Second, he attempts to defend what I call the traditional counterfactual approach to such an analysis, at least on two of the interpretations that I criticize in the book. I will consider the two lines of criticism in turn.

Eggleston claims that some of the implications of my contextualist account of "harm" are "quite strange." Indeed. If my contextualist account were the standard account and already widely accepted, I would not have bothered to include it in a book arguing for what I think is a radically different approach to some fundamental ethical issues. But the relevant strangeness supposedly goes beyond mere unfamiliarity to the point at which the implications are too strange to be acceptable. He illustrates this claim with an example in which my approach would allow him to say "yesterday I truthfully called an act a harming, but the act was not a harming" (for the detailed example, see Eggleston's paper).

Although his example roughly gets my position correct, there is some looseness of phrasing, which normally would not matter, but gets at a crucial distinction for my contextualist account (and many others). In the example, Eggleston uses the expression "I called the act a harming." This is ambiguous between (at least two) crucially different interpretations: (i) "I uttered the sentence 'the act was a harming'"; and (ii) "I said that the act was a harming." The difference between direct and indirect quotation is important here. It is clear from the example that Eggleston intends the first reading, since that is the only one on which the subsequent claims in the example could come out as true. But the strangeness, if any, attaches far more to the second reading than to the first. Compare:

> **A.** In a conversation yesterday, in full knowledge of the facts of the act and its consequences, I claimed that the act occurred here. That claim expressed a true proposition. But the act did not occur here.

And:

> **B.** In a conversation yesterday, in full knowledge of the facts of the act and its consequences, I uttered the sentence "the act occurred here." That utterance expressed a true proposition. But the act did not occur here.

It is easy to see how **A** is problematic (to say the least), whereas **B** need not be. In **A**, the referent of both occurrences of the indexical expression "here" is determined by the same context (the context in which **A** is uttered or written). In **B**, the referent of the first occurrence of "here" is determined by the context in which the directly quoted sentence was uttered, whereas the referent of the second occurrence of "here" is determined by the context in which the whole of **B** was uttered (or written). Given the possibility of different contexts and different referents, **B** is unproblematic. I have encountered many criticisms of the kind of contextualism in epistemology that my ethical contextualism parallels that rely on eliding the difference between **A** and **B**. Although Eggleston does not do that, the intuitive appeal of his example certainly seems to depend on such an elision.

But Eggleston does not consider the kinds of examples he produces to be merely "infelicitous." He adds that they "strongly appear to express false propositions, in the same way as statements such as 'Yesterday I said the act was done in Colorado, and that claim expressed a true proposition, but the act was not done in Colorado.'" This statement, of course, expresses a false proposition, because the referent of "Colorado" is not affected by the context of utterance. There is no relevant indexical element. So Eggleston is claiming that harm ascriptions likewise do not contain the kind of indexical element that I suggest they do. What arguments does he produce for this claim?

Eggleston points out that my account requires a speaker to possess certain kinds of information in order to be confident that their utterance of a sentence of the form "Act A harmed person P" expresses a true proposition. In particular, they need to know what the "appropriate alternative" is. But on a fairly straightforward interpretation of his own preferred account of harm, the speaker also needs to know what the appropriate alternative to the action is. It is just that the referent of "appropriate" is not affected by the conversational context of utterance. However, the speaker still needs to know what would have happened if the agent had not used their agency, either through inactivity or absence from the scene. This crucially depends on which of the two accounts

(inactivity or absence from the scene) is correct (more on that later). It is clear, then, that Eggleston objects to my account requiring knowledge of the conversational context in which the claims about harm are uttered. He illustrates this objection with an example in which such knowledge is required in order to tell whether a speaker's utterance on a previous occasion expressed a true proposition.

I have a couple of things to say about this objection. First, the kinds of cases which Eggleston presents would be, on my account, quite rare. Most harm claims will express the same propositions in most ordinary contexts, or at least propositions with the same truth values. If I poke Ben in the eye with a sharp stick, and you say, just afterward, "that act harmed Ben," you will almost certainly express a true proposition. Likewise, almost every conversational context in which my act is discussed will be such that "Alastair's poke harmed Ben" expresses a true proposition. We do not need to know all the details of an actual conversational context in order to be fairly confident that what is picked out as the "appropriate" alternative to poking Ben in the eye with a sharp stick is some act that leaves Ben better off than had he been so poked.

Second, it is not difficult to find other expressions, whose utterances similarly depend on conversational context to determine which propositions they express. Imagine a politician, who addresses two different audiences on two different days on her proposal to reintroduce public executions for theft. Suppose that she, in both venues, first describes her proposal, and then, in response to audience reaction, utters the sentence "my proposal is popular." The audience reaction on the first occasion is enthusiastic (unironic) cheers. What she is saying, then, is something like "my proposal is popular *with you*." In order to know whether she expresses a true proposition in the second venue, we would have to know at least something about *that* audience. If they are also enthusiastically in favor of bringing back public executions for theft, the utterance will express a true proposition. But if they are opposed to that proposal, it will not. Suppose the proposal in the second venue is met with mostly shocked gasps, in response to which the politician says "my proposal is popular," in a sarcastic tone of voice. Knowing these details about the different reactions would settle any doubts about whether the claims express true propositions. This example also illustrates my previous point, that actual uncertainty about which proposition a speaker expresses in a given context may be quite rare, and easy to settle.

Perhaps Eggleston will say "fine, I agree that there are other expressions which contain tacit indexical elements, such that there may be situations where we need to know details about conversational context in order to tell which proposition is being expressed. But my claim is that expressions concerning harm are not among them." And what is his argument for this? That my view "implies that people generally misunderstand the semantics of the word 'harm'. People generally think that the proper application of the word depends on what the act being considered consisted of, its effects, and so on – not the conversational context in which it is used." He claims that people know what they mean when they use "harm" in expressions such as "that act harmed me," and that the meaning does not involve an indexical element. Eggleston considers some of the examples I cite in my book of expressions that were formerly thought to pick out real entities (e.g., gods, phlogiston), but are now generally (phlogiston) or at least widely (gods) acknowledged to pick out nothing at all. He counters that "there is a significant difference between people being mistaken about what entities exist and people being mistaken about the basic functioning of accepted moral categories such as right and harm." He tries to further motivate his claim that attributions of

semantic error are less plausible in the case of harm ascriptions than in the other kinds of cases I put forward:

> Some kinds of widespread error are more plausible to assert than others… The claim that words such as 'harm' and 'benefit' have an indexical element would be more plausible if one could point to more specifically relevant examples, such as terms whose indexical elements used to be obscure to people for some reason, but that are now generally acknowledged. It is unclear whether such examples can be found.

In fact, I think it is remarkably easy to find such examples. Here is one: "sacrilegious." And here is another: "polite." It used to be understood, at least by many, that the standards for something counting as sacrilege were universal and unchanging. With more widespread skepticism about religion, and knowledge of diverse religious traditions, it is at least widely accepted that "sacrilegious" picks out different standards in different conversational contexts (which vary at least according to which religious traditions, if any, are accorded respect in that context). Mutatis mutandis for standards of etiquette.

My most serious objection to Eggleston's claim that my view involves an unacceptable attribution of semantic error to speakers is that it would seem to apply to his own view as well. His view is that to say that an act A harmed a person P is to say that the act resulted in the person being worse off than they would have been had the agent not exercised their agency, either through (i) being inactive at the time of the action, or (ii) being absent from the scene at the time of the action. He attempts to defend both accounts against my charge of giving counterintuitive results in my case of the button pusher who pushes button 99, with the result that a victim suffers terrible pain, when pushing any other button would have resulted in less pain (from almost as much, all the way down to none at all), but pushing no button would have resulted in a tiny bit more pain.

On both interpretations (i) and (ii) of the lack of agency version of the relevant counterfactuals, the agent benefits the victim, because no button would have been pressed, had the agent remained inactive or not been there in the first place. Eggleston's defense consists mostly of simply insisting on one or other of the two interpretations at issue ("but the victim would have been *even worse off* if the agent had remained inactive or not been there at all"), pointing out that to say that the action (of pressing 99) benefited the victim is not to give (much) moral credit to the agent, and emphasizing the moral reasons to press other buttons instead. Interestingly, this defense, if it were successful, would reinforce one of the points I want to make about the appropriate role for the concept of harm in fundamental moral thinking, which is, basically, none.

If we are forced to say that the agent does not harm the victim, but actually benefits her, in this example, because there is at least one other option on which the victim is even worse off, we have to admit that these facts about harm and benefit are completely irrelevant to how we, or the agent, should think about this particular choice. Once we have laid out all the options, from pressing no button all the way down to pressing button 0, we have all the information about the morality of the choice. We know, for example, that pressing 99 is the second worst thing that agent can do, only slightly better than not pressing any button, hugely worse than pressing 0, and involves no more effort or risk to the agent than pressing 0. That tells us all we need to know about the morality of the choice to press 99. It was pretty damn bad. It is not as if the addition of

the "fact" that pressing 99 did not actually harm the victim, but even benefited her (slightly) improves the morality of that choice, even a smidgen. On this account, the fact, if it is a fact, that pressing 99 was a benefit and not a harm is entirely morally epiphenomenal. This is important to Eggleston's treatment of the example. He is at pains to point out that classifying Agent's action of pressing 99 as a benefit, not a harm, in no way excuses the action. But here my contextualist account seems to have an advantage. While I agree that the notion of harm does not play a fundamental role in morality, my contextualist account *can* illustrate why we generally care about harm ascriptions. The features of a conversational context that help pick out which alternative is referred to as the "appropriate" one include our strong moral concerns. In the button pusher case, I claim that it would be very hard to imagine a context which did not pick out pressing 0 as the appropriate alternative, because it is hard to imagine a conversational context in which allowing unnecessary and easily preventable suffering is accepted. This is also why I am somewhat puzzled by Eggleston's variation of my case involving even more levels of possible suffering, brought about by pressing buttons with even higher numbers. In all such cases, pretty much every conversational context would still pick out pressing 0 as the appropriate alternative (it is still just as easy and risk-free as pressing any other button), so pressing 99 would still be correctly described as "harming" victim.

Return to Eggleston's criticism that my contextualist account of harm entails an unacceptable misunderstanding of the semantics of the word "harm." My account explains "Act A harms person P" as "A resulted in P being worse off than they would have been had the appropriate alternative been performed." This is a version of a counterfactual account, the general form of which is shared by many popular approaches to harm. Eggleston's preferred approaches are also versions of the counterfactual approach. Presumably it is the influence of conversational context that Eggleston claims is not generally understood. It is worth noting that standard counterfactual accounts are also affected by conversational context, though not necessarily in exactly the same way. And such accounts are also quite popular. But both Eggleston and I reject such accounts, so let us look at his preferred approach.

The two approaches that Eggleston defends are versions of the counterfactual approach where the antecedent of the counterfactual is fixed in a way that does not depend on conversational context. Instead, the antecedent picks out an alternative in which the agent is either (i) inactive at the time of the action (presumably immobile?), or (ii) absent from the scene at the time of the action. So, according to Eggleston, when I say that your act harmed me, I am saying either that your act resulted in me being worse off than I would have been had you remained inactive at the time of the action, or that it resulted in me being worse off than I would have been had you been absent from the scene at the time of the action.

Given the number of different accounts of harm that have been proposed and defended in recent years, both variations of the counterfactual approach and other, non-comparative, accounts, it is overwhelmingly likely that no particular account is accepted by most philosophers (and the same probably goes for most ordinary speakers of English). So, any account of harm entails that many, probably most, people misunderstand the semantics of the word "harm." It is not at all clear why the level or type of misunderstanding that my account would attribute to those who do not accept it (yet) should count against it any more than the level or type that either of the accounts Eggleston favors would attribute to those who do not accept them.

Furthermore, the two accounts that Eggleston defends against my criticisms differ from each other crucially, so at most, one of them can be correct. To see this, consider again my example of button pusher and the 100 buttons. Add the detail that a different person, Back Up, is in the vicinity, and only wanders away when they see Agent at the control panel. If Agent had not been there in the first place, Back-Up would have been given the choice, and would have pressed button 0. Now we get different results from the two versions. On the inactivity version, we still get the result that Agent benefitted Victim, because if Agent had remained inactive, but present, Victim would have suffered at level 100. But on the absence from the scene version, we get the result that Agent harmed Victim, because if Agent had not been there in the first place, Back-Up would have been given the choice, and would have pressed button 0. My point in raising this is not to argue for one over the other. I reject both. But it is clear that Eggleston would have to reject at least one of these two accounts, and thus attribute the relevant semantic misunderstanding to those who accept the other, as well as to all those who accept different accounts altogether.

Sinnott-Armstrong

Walter Sinnott-Armstrong raises many interesting, and challenging, issues for the contextualist approach that I sketch in the book. I do not have space here to address any of them with the care and detail that they deserve. I will, however, attempt to at least gesture in the direction of answers to some of the more central issues, and to clear up some confusions.

Walter presses me to be clearer about what, exactly, my version of contextualism is supposed to say. In introducing my analyses, **G-con**, **H-con**, and **R-con**, Walter points out that the contextualism I am appealing to is not explicitly invoked in the analysis, and so has to be inferred from my discussion. Likewise, my claim that "appropriate" functions as an indexical only introduces context to the right side of my analyses. Walter clears up any potential confusion by suggesting meta-linguistic versions of my analyses. He is correct to suggest that it is the meta-linguistic versions that I am exploring. I had thought that the simpler formulations I included in the book, with the accompanying discussion about context and indexicals, would be enough to avoid confusion, but I should have been clearer that my contextualist proposal is that seemingly non-contrastive sentences (e.g., "my action was good," "your action harmed me," etc.) can be used to express contrastive propositions, and that the context of utterance (partly) determines which contrastive propositions are expressed.

I will not, however, go along with Walter's suggestion that I am talking not about specific alternatives, but rather sets of alternatives. In each of the formulations I give, there is a reference to "the appropriate alternative." I do really mean to specify, in each case, a single alternative. The reason is that the version of contextualism I am exploring has the context determine the *standard* that has to be met for an action, in that context, to be classified as right, good, harmful, beneficial, or the like. All the options in a given choice situation could be ranked against each other, in terms of the overall value of the worlds in which they are performed. An action, A, that is at least as good as a particular option, X (which may be A itself), will also be at least as good as any option that is no better than X.

Take the suggested account of "right":

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

The idea is that, in any given conversational context in which an action is being discussed and assessed, there may be at least a rough agreement (probably tacit) about just how good (comparatively) an action must be in order to be judged "right." In order to draw out such a standard, we could imagine asking something like "what would it take to act rightly in that situation?." The answer could be of the form "you've got to at least do X, certainly nothing worse than that." Sometimes, of course, the different parties to a conversation may actually have different standards in mind, but fail to realize it, because the action in question either meets, or fails to meet, all of the different standards.

Walter's most challenging comments concern the need for me to say more about how context determines which alternatives are appropriate. The problem appears to be that I do not provide even general guidelines for how different conversational contexts determine the relevant standards for the different ethical judgments that are made in those contexts. In the absence of such guidelines, we will not be able to apply the different accounts, in order to determine whether true or false propositions have been asserted:

> The problem is that reasonable people could easily disagree about which alternatives are or are not appropriate, so Norcross needs to say not only which alternatives are appropriate in a context but also why those alternatives are appropriate in that context but not in other contexts.

Before I address this specific objection, I should distinguish at least three different ways in which people might appear to disagree about the moral character of an action being discussed in a context, even while operating from within a consequentialist perspective. First, they might disagree about what the actual consequences of the action were, and what the consequences of specific alternatives would have been. This kind of factual empirical disagreement, while challenging for the practice of practical reason, raises no problems for ethical theory. Uncertainty about the consequences of our actions is simply part of the human condition.

Second, people might agree about the empirical facts, but disagree about the values involved. For example, suppose we are discussing a case in which a dying person asks their long-term partner whether they have ever cheated on them. The partner did, indeed, cheat many years earlier, and has to decide whether to tell the truth or lie to their dying partner. We might agree that telling the truth in this situation would lead to more net unhappiness, but also more accurate beliefs. We may disagree, however, on the net comparative value of the two alternatives. Perhaps I think that only hedonic subjective states have intrinsic value, and that the alternative with the comforting lie would be better overall than the alternative with the painful truth. You, however, think true beliefs also have intrinsic value (or more specifically, that the truth of a belief contributes positively to its intrinsic value), or that honesty has intrinsic value, or that the satisfaction of a desire to know the truth has intrinsic value (or some combination of these things), and that the extra intrinsic value added by some combination of these factors would more than outweigh the hedonic loss involved in telling the truth, and thus that the alternative with the painful truth would be overall better than the alternative with the comforting lie.

Such axiological disagreements are also fairly common, though probably less common than straight empirical disagreements. Clearly, this is not the kind of disagreement Walter has in mind. But it still raises tricky questions. How do we settle the question of

which contrastive sentence expresses a truth in that situation? "Lying has better results than telling the truth," or "telling the truth has better results than lying." Presumably, the truth hinges on the axiological question of what does have intrinsic value. While there is, presumably, a fact of the matter, there is certainly no easy way to discern that truth, as the history of axiological disputes demonstrates. Such disputes are inevitable. The (current) lack of a method to settle them does not threaten the plausibility of a consequentialist contrastive approach that analyzes "better" in terms of overall net comparative value.

So what of the kind of disagreement that Walter is raising as a problem for my approach? This is the third kind I have in mind. Suppose that two people are discussing a situation in which an agent rescued seven people from a burning building, but could have saved up to ten people, with escalating degrees of difficulty and self-sacrifice. Both people agree on all the comparative judgments. Saving ten would have been better than saving nine, which would have been better than saving eight, and so on. And they agree on all their judgments of just how much better or worse each alternative would have been than all the others. But one of the evaluators claims that Agent's act was good "because rescuing ten is not an appropriate alternative," but the other claims that Agent's act was bad "because rescuing ten is an appropriate alternative." Do we need to settle this dispute, in order to see which non-contrastive sentence expresses a truth in this context?

Suppose, in the above case, that the first speaker declares that Agent's act was good, because she assumes that rescuing any number of people from the burning building, given the level of risk and self-sacrifice involved, is at least as much as can be reasonably expected. If pressed to justify her claim that the act was good, she is more likely to talk about "reasonable expectations in the circumstances" than about "appropriate alternatives." My use of the term is not intended to echo what people would actually say, if asked to explain their judgments. (I should add, at this point, that I now prefer to talk of "appropriate options," rather than "appropriate alternatives," because "alternative" suggests something different from what was actually done, and I want to leave open that the action in question is itself the appropriate option.) Now suppose that the second speaker declares that she disagrees, and claims that Agent's act was bad, because she assumes that rescuing all ten was what could be reasonably expected in the circumstances. This disagreement, as Walter correctly points out, illustrates how "the speaker's context, the audience's context, and the agent's context can come apart." As he says, I am definitely not talking about the agent's context, which would involve the kind of invariantism I am denying.

But what of the difference between speaker context and audience context, which may be apparent even in conversations taking place in person? This seems to be a case of speaker meaning and hearer meaning coming apart. This is not uncommon. Suppose I say that Steve Nash is tall, and you appear to disagree. We both agree that he is six feet three inches tall. Perhaps we appear to disagree, because I mean by "tall" something like "significantly above the expected height for a Canadian adult male," but you would mean something like "significantly above the minimum expected height for an NBA basketball player." Once we explain what we mean, you agree that Steve Nash is tall in my sense, and I agree that he is not tall in your sense. But the burning building case seems different. The two speakers disagree about what is reasonably expected in the circumstances. But we can imagine a different height disagreement that mirrors this too. Suppose I say that Lebron James is tall, and you appear to disagree. This time, I justify my claim by saying that James is significantly above the minimum

expected height for a National Basketball Association (NBA) player. You say that he is not, perhaps because you disagree about what the minimum expected height is, or what is takes to be "significantly" above that height (or both). This seems to be a pretty common, and unproblematic, case of speaker meaning and hearer meaning coming apart. Do we need to settle the dispute about what the expected minimum height is for an NBA player, or what it takes to "significantly" exceed that height, in order to tell which non-contrastive sentence about height expresses a truth in this context? Why not simply say that the proposition expressed according to the speaker meaning is true, but the proposition that would have been expressed according to the hearer meaning was false?

In many contexts, speaker meaning and hearer meaning will be close enough that both speaker and hearer will agree on the truth value of the propositions being expressed or understood. But sometimes, they will come apart. Likewise, in my proposed contextualism about moral terms, there may often be enough (probably tacit) agreement about what could be reasonably expected of the behavior of the agent in question in order to live up to (or down to) the standard being invoked (goodness, rightness, supererogation, harm, etc.), that the proposition intended to be expressed by the speaker will have the same truth value as that understood by the hearer(s), and may be the same proposition. But in situations where the expectations that set the standards differ enough between speaker and hearer (and maybe even between different hearers), the proposition intended to be expressed and the proposition(s) understood to have been expressed may be different enough to have different truth values. I do not see this possibility as sufficiently different from many other aspects of linguistic communication to pose a special problem for my approach. Successful communication depends on a certain amount of shared understanding, expectation, and so on. Just how much needs to be shared is often a vague matter, depending partly on the aims of the conversation. You and I can have many successful conversations about chairs, for example, while differing in our understanding of just what is required in order for something to be a chair. We may agree, for example, that my chair is more comfortable than yours, but that yours is bigger than mine. We may only discover that we have slightly different understandings of "chair" when your enthusiastic recommendation of certain objects in a camping store as "chairs" is met by blank stares from me (because I would classify them as "stools").

But why not simply avoid these issues altogether, and endorse Walter's contrastivist approach? This is how he puts the central idea: "the sentence-meaning of 'The action is good' is simply 'The action is good in contrast with the relevant alternatives.'" I must admit that I am struggling to see just how different this is from my own account (as he says, we are not really that far apart). We agree that the fundamental moral facts are all comparative. "Good in contrast with the relevant alternatives" seems to mean pretty much the same thing as "better than the appropriate alternative" (setting aside the issue of whether we are talking about a single alternative or a class of alternatives). Both "appropriate" and "relevant" function as indexicals, and I would be happy to substitute "relevant" for "appropriate" in my accounts. Which alternatives are picked out as "appropriate" or "relevant" vary according to the context of utterance. I do slightly prefer "appropriate," though, because I think that normative expectations often play a role in determining a linguistic context. When Walter describes the speaker meaning of an utterance in a context as what the speaker "intends to convey," he seems to be talking about the same thing as what I am talking about when I say that moral terms are used to express different propositions in different contexts.

Perhaps the main reason for me to prefer my account is that the role of conversational context emphasizes the social cooperative nature of communication. What a speaker "intends to convey" by an utterance in a context is not entirely up to them. Their understanding of the shared beliefs and expectations in their linguistic context will at least influence what they can intend to convey. Their understanding may, of course, be deficient, perhaps severely so, and what they intend to convey (what Walter calls the "speaker meaning") may not be conveyed at all. Their audience may incorrectly believe that the speaker intended to convey one thing, when they in fact intended to convey something quite different. Or, to put it in my terminology, the proposition the speaker intended to express may be different from the one the audience took the speaker to have expressed. We may wonder, in this situation, whether the speaker actually expressed one proposition, but the audience mistakenly believed that they expressed a different proposition, or rather the speaker intended, but failed, to express a particular proposition, because their understanding of the conversational context was deficient. I admit that I do not have a settled view on this matter. One advantage of Walter's talk of the speaker intending to convey something is that it sidesteps this question.

Mason

Elinor Mason wants to suggest something like an account of wrongness that connects with blameworthiness:

> what we need to look for is an account of wrongness that connects it to blameworthiness – the aptness, or fittingness of blame – where the agent's intentions – the quality of her will – is what makes blame apt or fitting.

She also refers to my discussion of Mill's famous suggestion regarding wrongness and punishment/blame:

> "We do not call anything wrong, unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow-creatures; if not by opinion, by the reproaches of his own conscience."[1] The point here is *not* that something is wrong if we ought to punish it, or it would be optimific to punish it (as Norcross imagines the options). It is that there are demands such that failure to comply makes punishment fitting.

The topic of blameworthiness (whether connected to other forms of punishment or not) is clearly an important one, that I do not discuss much in the book. In fact, early on I say this:

> For a consequentialist, the only consideration of direct relevance to the moral evaluation of an action (as opposed to the character, blameworthiness, etc. of the agent) is the comparative value of its consequences. (6)

This, of course, leaves open the possibility of something like blameworthiness being of important *indirect* relevance to the moral evaluation of an action. I take Mason to be

---

[1]Mill, *Utilitarianism*, ch. 5, para. 14. Quoted by Norcross on p. 29.

urging me to consider views on which blameworthiness is of direct relevance to at least some moral evaluations of actions. I accept that I should say more about how blameworthiness figures into moral evaluations of actions, as well as of agents, but I am inclined to stick to my guns here in denying it *direct* relevance to act evaluations.

In discussing Frank Jackson's famous example of the doctor with three options, one of which will completely cure a patient, another of which will kill the patient, and the third of which will ameliorate the symptoms, but not cure the patient, Mason urges a form of subjectivism about obligation, that focuses on the agent's position and her intentions. The connection she wants to make between the subjective states of the agent and rightness or wrongness concerns what she calls the "quality of an agent's will." She also invokes the notion of an agent's intentions making blame "apt" or "fitting." This is clearly supposed to be something different from an evaluation of the consequences of blaming or punishing an agent who has certain intentions, or qualities of will. But what, exactly, is it for blame or punishment to be an apt or fitting response to an agent's behavior?

Richard Arneson (in an AMC session on my book at the APA) singled out "failing to show due impartial consideration for persons, or other sentient beings," as making blame apt. I agree that the fact that someone displays this failing is a moral defect. It is certainly a morally relevant negative fact about the agent. We could, if we wanted, decide to employ the term "wrong" to connect with such a fact in various ways. But this is all consistent with my claim that such subjective features of an agent are of *indirect* relevance to the moral status of their actions. Likewise, when thinking about what it is that makes reactions like blame "fitting" or "apt" when applied to an agent's intentions (or other features that we think are related to the "quality of will"), we can compare the reasons behind internalizing (and encouraging others to internalize) certain practices of assigning praise and blame. It may well be that a practice of internalizing the practice of blaming and labeling as "wrong" the same actions is supported by comparatively stronger reasons than any practice that separates these two activities, and so we can construct a fruitful practice connecting blame with wrongness.

Further, we can compare different such practices (using different practical standards for the assignment of blame and wrongness), and feed them into judgments about fittingness. Again, all this is consistent with my claim that the only thing of *direct* relevance to the moral status of an action is the comparative value of its consequences. The reason why I would not go further, and incorporate notions like the aptness or fittingness of responses to certain aspects of an agent's mental states into a more ground-level or fundamental judgment about an action's moral status is my general suspicion of placing much weight on the kind of gut-level intuitions that seem to drive such fittingness judgments. Given how vulnerable our moral intuitive reactions are to perverse societal influences, how much sway those in power (religions, kings, rulers, corporations, etc.) have over our moral emotional responses, I try to avoid using all but the most general and universal of such judgments (e.g., pain is bad, pleasure is good) in *constructive* theory building (as opposed to appealing to the moral emotions of my interlocutors to show some deficiency, such as internal incoherence, in their own approaches).

But cannot we ground judgments of blameworthiness in some kind of inconsistency between an individual's character (perhaps as displayed in their intentions, or other aspects of the quality of their will) and the kind of consequentialist normative theory I defend? This would not necessarily require too much appeal to gut-level intuitions about particular persons or actions. We could, perhaps, identify certain negative

features of a person's character, using reasons internal to the version of consequentialism we are embracing. Those features that displayed the greatest tension with the central aspects of our theory (such as a desire to increase suffering in other sentient beings, a tendency to give vastly unequal consideration to the interests of different sentient beings, a callous indifference to the suffering of others, etc.) could be picked out as "apt" for, or "deserving of" blame. We could also identify actions that characteristically stem from such "blameworthy" elements of character, and derivatively extend the notion of blameworthiness to those actions. Finally, we could defend the practice of labeling such actions as "wrong," and of "demanding" that they not be done, by comparing specific versions of such practices with alternative practices in terms of overall social utility. All of this is quite consistent with my claim that neither wrongness nor demands are fundamental moral features of actions.

Mason suggests that moral reasons without demands loses blameworthiness, because

> An agent is only blameworthy for failing to abide by reasons if those reasons command… Not following a reason that we are not commanded to follow cannot be something we are criticisable for. Being imprudent (or irrational, or immoral on Norcross's view) is just like failing to abide by trivial reasons.

But whether it is or not depends on the consequences of adopting such practices. If, as seems likely, societies that adopt blaming practices that differentiate between kinds of reasons are better off, other things being equal, than societies that do not, then these things are not "just like" each other. As an illustration of this point, consider Mason's contrast between the case of the doctor advising parents about potential life-saving treatment for their child, and advising an adult about comparable treatment for themself. In the latter case, the doctor's role is to advise, providing the adult patient with the information they need to make their own choice. If the adult chooses to forgo treatment, despite the alternative being almost certainly worse for them, the doctor cannot force the treatment on them. She can only provide the reasons. But in the case of the child, the situation is different:

> An adult patient in full possession of her faculties can refuse treatment, even if the reasons point to treatment… But in the case where the patient is a child, we think that the reasons do more than point, they *demand*. The doctor *must* do what she can to save the life. The parent *must not* interfere. We do not think of the law that allows the medical establishment to overrule a parent on this as merely convenient or expedient. It's the morally right law to have.
> In the case of an adult patient,… if she chooses to do what is in her own best interest, there may be no sense of mustness, of obligation, about it. It is just the choice she makes. But other choices she makes are not optional in this way. Some reasons command.

Mason is here appealing to powerful intuitions. But I do not think we need to follow them to her destination. We can account for, and accommodate, such intuitions within the framework I defend. Mason says that the case of the child goes beyond there being good reasons to save the child's life. Maybe. But does it go beyond the reasons to save the child's life being immensely stronger than those to let the parents let the child die? She says that we do not think that the law that allows the doctor to override the parents is "merely convenient or expedient," but that it is "the morally right law to have." But

there is nothing necessarily "mere" about the level of expedience that applies to the justification of such a law. It is quite plausible to suggest that it is a morally far better law to have than the alternative that allows parents to let their children die in such situations. If the difference in strength of reasons supporting these alternative legal arrangements is large enough, as seems quite likely, the greater reason can easily take on a psychologically insistent character. When Mason declares that "some reasons command," she may be doing no more than alerting us to the undeniable fact that some reasons are far more insistent than others. And the reasons we have to encourage thinking about adult self-interested choices differently from choices made by adults about children are comparatively strong.

Part of my disagreement with Mason, then, might be simply terminological. She may well be using "demand" in a more expansive sense than I am. She says:

> If, however, we are clear that there are no demands, we can only point out that a choice was immoral, and that term has no inherent censure.

But "immoral," as I understand it, is as inherently censorious a term as it gets. When I say that there are no demands, at the fundamental level of moral theory, I am not saying that there are only descriptive claims to be made about choices. The claim that one choice is much better supported by moral reasons than another choice is a normative claim, and the judgment that a choice was vastly morally inferior to an alternative is inherently censorious. I am simply claiming that once such a judgment has been made, at least at the level of fundamental ethical theory, there is no *further* fact of the form that morality "demands" that the alternative be made instead (though the *practice* of talking about the "demands" of morality may well be better supported by moral reasons than many alternative practices). Perhaps, then, I really am talking about demands, at least in the sense that Mason understands the term. Maybe, so long as it does not incorporate a notion of blameworthiness that goes beyond what I have already sketched.

Kagan

I must admit that I have the most difficulty knowing what to say in response to Shelly. The main reason is that I thought I had been quite clear in laying out what my position is, but Shelly accuses me of a singular lack of clarity. As I see it, his criticism seems to be that I am never clear enough about whether the scalar approach that I advocate is one that "does without" rightness (and the other deontic categories I discuss), or rather is one that combines a scalar approach at the fundamental level with a contextualist account of rightness. Only the former, Shelly seems to be saying, would be a truly radical theory. The latter would be just another version of consequentialism to consider alongside maximizing and satisficing versions. I will try to address what I see as a fundamental confusion in this waying of putting things, but I am not confident that what I say will be satisfying, at least to Shelly.

So, am I just advocating another account of rightness, to be evaluated alongside a maximizing account and a satisficing account? The reason to think I might be doing that seems to be that we can take what I call "Core Consequentialism," and simply tack on a maximizing or satisficing account of rightness, instead of the contextualist account of rightness that I suggest as the most promising alternative to simple eliminativist talk. As a reminder, this is what I call "Core Consequentialism":

> **Core Consequentialism (CC):** An action is morally better or worse than available alternatives, and *thus* there is greater or lesser reason to opt for it, *entirely* to the extent that the world containing it is overall better or worse (contains more or less net intrinsic value) than the worlds containing the alternatives. (Emphases not in the original)

Shelly suggests that we can see different consequentialist theories as the combination of CC with various accounts of rightness (and maybe other moral properties too). On reflection, I can see why he might think this. I was not really clear enough in explaining my position. A lot turns on one of the words which I have emphasized in the above quote: *entirely*. Likewise, when I say in a few places that scalar consequentialism has no room for rightness (and the like) at the *fundamental* level of the theory, I am trying (though apparently not entirely successfully) to highlight one of the important differences between my scalar approach and traditional maximizing and satisficing approaches.

The proponents of maximizing and satisficing versions of consequentialism can be seen as advancing competing accounts of the property of rightness. They agree that there is such a property, but disagree about the correct account. The maximizers claim that rightness (of actions) is the property of maximizing net value, whereas satisficers claim that rightness is the property of leading to "enough" net value. But both maximizers and satisficers agree that they are talking about the same thing: the property of rightness. Further, they (if I have understood them correctly) agree that the property of rightness has basic, or fundamental, moral significance.

The contextualism I explore, and at least tentatively endorse, is not a contextualist account of the property of rightness. My claim is that there is no such thing as the property of rightness. There are many different properties, which are picked out by different utterances involving the word "right" (and related terms). The contextualist approach I explore is a linguistic account of the use of "right" (and related terms). This account entails that different utterances (or other tokens) of "right" pick out different properties, depending on features of the linguistic context in which they are made. In one context, the property picked out may be that of being no worse than (in terms of net consequential value) the option of lying beside a shallow pond discussing Ayn Rand, while a bunch of children drown (but not actively throwing children to their watery deaths). In another context, the property may be that of maximizing net value.

That a linguistic context selects a particularly permissive property as the one picked out by "right" or "permissible" may itself be assessed as better or worse than alternative contextual features. To suggest that one Randian may have expressed a true proposition when he said to another Randian "it's perfectly permissible for us to continue our discussion of Atlas Shrugged while all those children drown," is in no way to endorse either the behavior described or the act of describing it. It is context-independently true both that it would have been much better had the Randians abandoned their book discussion and saved some lives, and that it would have been much better had they been inclined to assert sincerely "it's wrong of us to sit here and watch the children drown."

So, what does this tell us about the *significance* of "right," "good" (as predicated of actions), "harmful" (again, predicated of actions), and so on? I do not think I can put it any better than I do in the closing sentences of my book, so I will quote myself here:

> 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, or harmful varies according to the context in which it is made, then rightness, or goodness, or harm can no more be properties of actions themselves than thisness or hereness 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. Which properties we are interested in will vary from context to context. But we are not mere passive observers of, and adherents to, conversational contexts. We can also play an active role in shaping those contexts. In fact, a fruitful avenue for promoting the good is the very activity of context shaping. (151–52)

I am not simply suggesting an alternative consequentialist account of rightness, to be evaluated alongside maximizing and satisficing consequentialism (and any others). I am arguing that there is no such thing as <u>the</u> property of rightness, and *a fortiori* that there is no such property that has fundamental moral significance. As Shelly says, I give various arguments against the claim that the property of maximizing net goodness, in particular, is fit to play the role of <u>the</u> property of rightness with fundamental moral significance in a consequentialist theory. I do not have the space here to fully assess everything he says about my arguments, but there are two points I want to clarify.

First, in discussing what I say against a particular argument for a maximizing account of rightness, Shelly says:

> The other account, the one I want to focus on, holds that "right" just means "supported by the strongest (moral) reasons" or "what we have most (moral) reason to do" (34). Norcross notes that if this were right, then the maximizing account of rightness would hold as a trivial conceptual matter, so that anyone who believes that some acts are morally superior but not obligatory must be "confused about the meaning of words" (35) – which is a highly implausible thing to claim.

In response, Shelly says:

> But why should someone who accepts a maximizing version of consequentialism have to agree that the maximizing claim (that one is required to act in an optimal fashion) holds as a matter of the very meaning of the word "right" (or "required" or "permissible" and so on)? Why can't the maximizing claim simply be put forward instead as a substantive moral truth? As far as I can see, maximizing consequentialists can readily agree that the word "right" (for example) does not mean "supported by the strongest reasons" – all the while insisting, for all that, that an act is right if and only if it is supported by the strongest reasons.

I agree, of course, but I am puzzled as to why Shelly puts this forth as an answer to my criticism. I am criticizing the claim that "right" just means "supported by the strongest reason." This is a claim that has been advanced by some consequentialists. Shelly points out that we do not have to say that "right" just means "supported by the strongest reasons." Well, good. We not only do not have to say that, but we should not say that (speaking with the vulgar in my use of "should not"). So we agree. Perhaps Shelly's

point is that the linguistic argument I am criticizing does not even merit such scant attention. Perhaps he is right about that. But I have learned, over the years, to temper my confidence in my judgments of what positions or arguments are too implausible or feeble to merit serious discussion.

The second point of clarification concerns what Shelly says about my arguments that rightness, on either a maximizing or satisficing approach, does not provide any kind of reason for choice over and above what is already provided by the comparative net consequential value of the options being considered. Apparently, Shelly finds my example involving Smith and Jones confusing, but he modifies it appropriately, so I will consider what he says about the modified example. In this example, we are to imagine that we can either choose to make a difference to how much good Smith will do or to how much good Jones will do. By hypothesis, Smith is deliberating between doing something that would be right, according to whichever consequentialist account of rightness we are considering, and doing something that would be worse than this by enough to be not right. Jones, on the other hand, is deliberating between two options, both of which fall short of the relevant standard of rightness. Furthermore, the positive difference in net goodness that would occur if Smith chooses the better of her two options is the same as the difference if Jones chooses the better of his two options. My claim is that no consequentialist should take the supposed fact that Smith will perform a right action, if she chooses the better of her two options, but Jones will fail to perform a right action either way, as a reason to choose to persuade Smith to choose the better of her two options, rather than to persuade Jones to choose the better of his two options. Shelly points out that we can accept that the existence of an extra right action, performed by Smith, does not provide *me* with a reason to focus my attention on Smith rather than Jones, without accepting that it does not provide *Smith* herself with an extra reason.

Shelly is correct to point out that this particular example only supports the claim that the rightness, or not, of *someone else's action* does not have significance for an agent. But I address this very worry immediately after giving the argument Shelly is discussing. Here is what I have to say about it:

> Suppose that you, a maximizing utilitarian, wake up one day with a wicked hangover. You remember that yesterday you were grappling with two decisions. In the first one, you were inclined to perform act A, but were considering B instead, and in the second you were inclined to perform act C, but were considering D instead. In the first decision, B was better than A by 10 (net) hedons, and was the best of all your available options. In the second decision, D was better than C by 11 (net) hedons, but was actually second best to act E of all your available alternatives (but you had definitely decided against E). You remember that a friend managed to talk you up to the better of the two options in one of the decisions but not in the other. But you don't remember which it was. In thinking about this, qua utilitarian, you should clearly hope that you did D instead of C, rather than B instead of A. Even though this means that in neither decision did you do the right thing. The situation would remain the same, even if the difference between C and D were only 10.1 (net) hedons, or 10.01. So we are left with the bare claim that the difference between right and wrong is itself morally significant, even though it is irrelevant to decisions about what to do, and irrelevant to the value of states of affairs, and thus irrelevant to what a utilitarian, qua utilitarian, cares about or hopes for. If, despite accepting all this, a utilitarian were to insist that the difference between right and wrong is still "significant", the appropriate response would be to quote the

character Inigo Montoya, from the movie The Princess Bride. "You keep using that word. I do not think it means what you think it means." (24–25)

Inigo Montoya, played by the wonderful Mandy Patinkin, addresses that line to Vizzini, played by the equally wonderful Wallace Shawn, after Vizzini repeatedly describes as "inconceivable" something that has actually happened, or is happening. Although he does not specifically address this argument, Shelly does attempt to avoid the Princess Bride conclusion with the following account of "significance":

> But even if it is true that from the consequentialist perspective an act's having the property of being required is not in itself a reason to perform that act, it doesn't follow that the property is not a morally significant one. Knowing that an act is required may not give you an extra reason to do it, but it does tell you something crucially important about the nature of the reasons you have for doing it. In particular, being morally required is a property that an act has when the various reasons to perform that act are jointly (morally) decisive, settling the question of how someone who perfectly conforms to morality would act.

But, as I have said, I claim that there simply is no such property as the property of being required. So, it is not possible to know that an act is required, any more than one can know that the Easter Bunny has left chocolate for our consumption. If we adopt my contextualist linguistic approach, one can know that an act has the property that would be picked out by an utterance of "that act is required" in one particular linguistic context, and maybe also know that the same act lacks the property that would be picked out by an utterance of "that act is required" in a different linguistic context. Am I denying that there is a univocal property of acts that coincides with (or maybe just *is*) the property of being such that the various reasons to perform the act are jointly (morally) decisive? Yes. But why not just adopt what is suggested by the last line of the quote from Shelly's paper, "settling the question of how someone who perfectly conforms to morality would act"? Why not just say that being required is that property of acts that all those who perfectly conform to morality would perform, and which explains why such acts perfectly conform to morality? Well, what is it to perfectly conform to morality? Is it to act on the strongest balance of moral reasons? In which case, all my arguments against maximization reappear at this level. If not, what is it? I suspect that the curse of Vizzini is not so easily avoided.