

“What Should We Say about Contrived ‘Self-Defense’ Defenses?”

**Dan Farrell
Department of Philosophy
Ohio State University**

As a paradigm of the kinds of cases that raise the problem of the so-called “*Actio Libera in Causa*,” we might well begin with the case described by Finkelstein and Katz, of the defendant who “deliberately drinks himself into a state of irresponsibility in order to commit a crime while in that state.”¹ One central question such a case raises, Finkelstein and Katz note, is “whether we should deny the defendant a defense we would otherwise grant because the defense is claimed in the context of a plan to produce [that defense].”²

Suppose we are inclined to deny a defense in such a case and are inclined to deny it precisely because of the defendant’s machinations: such a case is quite different, we might say, from a case in which someone has been unknowingly drugged into irresponsibility and commits a crime while in that state. To allow a defense in the contrived case is to make a mockery both of the law generally and of the rationale for allowing a defense in the case of the defendant who was unwittingly drugged.

Much more needs to be said about this line of thought, and I’ll return to try to say a bit more about it near the end of my presentation. My real interest here, though, is in a rather different sort of case, to a discussion of which I’ll devote most of the presentation.

I have in mind a case in which someone deliberately provokes another person to attack him, in order to be able to harm the latter in self-defense and then use “self-defense” as a legal defense for what he’s done. I’m not sure how widely this sort of case would be recognized as an example of an “*actio libera in causa*”—Finkelstein and Katz give such a case as an example of several sorts of cases to which the “*actio*” doctrine has been “extended”—but, as I’ll try to show, thinking about such cases in the context

of the “*actio*” problem is both interesting in its own right and also potentially useful in helping us decide what to say about cases like the paradigm case with which we began.

First, though, some preliminaries. I’m not trained in the law, and the remarks that follow are not about what we should say *in the law* about the various kinds of cases I discuss. I’ll be talking about what we might be inclined to say from a *moral* point of view about such cases. (What I mean by this will become clearer in a moment.) I’m hopeful that discussing such cases from a general moral point of view will be helpful to those thinking about how to treat such cases in the law, but, apart from some very sketchy and very tentative remarks at the end of the paper, I shall not attempt to defend any legal applications in the present paper.

As a second preliminary point, I need to note a peculiarity of how I shall proceed in what follows. Over the past several years, I have published a series of papers defending a certain way of thinking about standard cases of self-defense and about the idea of being justified, morally, in defending oneself--with violence, if necessary--in such cases. In what follows I want to look at what we might call “contrived” (and hence non-standard) cases of self-defense from the standpoint of my own views about self-defense and its justification in standard cases, views that were developed without any thought about contrived cases and were meant to apply only to what I’ve just called “standard” cases: namely, cases in which one person knowingly and wrongfully attacks another person who bears no responsibility whatsoever for the former’s attack.

Briefly, my view about self-defense and its justification, at least in certain kinds of standard cases, is that, when it is morally justifiable, harm to another in direct self-defense is justified as a form of *distributive justice*: when, as in a standard case of an unprovoked attack, someone has wrongfully made it necessary that either he or some other (innocent) person must suffer some degree of harm, the latter is justified, as a matter of basic justice, in taking steps to see to it that the former is harmed, rather than

that she is harmed, at least if the harm thereby visited on the former is not radically disproportionate to the harm that would otherwise be suffered by the latter.³ This view, which of course needs both clarification and a defense, can be summarized a bit more perspicuously as follows:

P1: When someone knowingly brings it about, through his own wrongful conduct, that someone else must choose either to harm him or to be harmed herself, justice allows the latter to choose that the former shall be harmed, rather than that she shall be harmed, at least if the harm inflicted on the former is not radically disproportionate to the harm that would otherwise be inflicted on the latter.⁴

I cannot discuss this principle at any great length here, much less defend it in a systematic way, which is something I've done elsewhere. A few remarks may be helpful, however, before I ask my listeners simply to *assume* the principle's plausibility, for the sake of argument, and then turn with me to the cases that actually interest us on the present occasion: namely, non-standard, "contrived" cases of self-defense in which someone has deliberately provoked another person in order to give himself the opportunity to harm the latter and then defend his doing so as a matter of legitimate self-defense.

Notice first, then, that P1 is not based on utilitarian or any other consequentialist considerations, though of course a principle with identical content could be. Rather, as I conceive of it and use it, the principle is a *deontological* principle, alleging that defending oneself with harm to another in the relevant situations is a matter of *justice or fairness*—a matter of having a right to act in a certain way, that is, due to considerations of what justice or fairness allows—rather than a matter of what's best overall in some consequentialist sense.⁵

Secondly, it should be clear that P1 assumes a world in which a form of free will exists which is such as to support a very robust notion of moral responsibility, such that when someone freely and knowingly chooses a certain course of conduct, she or he can at least sometimes be held responsible for the foreseeable consequences of that conduct, all other things being equal. (No assumption is made, one way or the other, about whether free will and moral responsibility, of the sort I assume, require the

falsity of what is sometimes called “the thesis of universal causal determinism.” I aim to be neutral, that is to say, on the metaphysical issue of “compatibilism” versus “incompatibilism” in the free-will-and-determinism debate.)

Note, thirdly, the importance, given my deontological framework, of the proportionality requirement that P1 includes. This is not the place for a detailed discussion of what this requirement demands, but, very briefly, and admittedly very roughly, it requires that the harm inflicted in defending oneself, in accordance with P1, not be radically disproportionate to the harms one can reasonably expect to suffer if one does not defend oneself by harming the attacker in order to stop him from harming oneself. This requirement, like the right the principle supports, is, on my view, a requirement of justice (or fairness, which for present purposes I take to be synonymous with justice).

Fourthly, it should be noted that I am aware that there are all sorts of epistemic questions that real-world applications of P1 would have to face and hence that a defense of P1 would have to address: problems like the degree of certainty (that one must defend oneself or be harmed) required for appeal to P1, the likelihood that one’s reactive force will be effective rather than ineffective or even make the situation worse, and so on. These are important issues—and there are plenty of others—for someone who accepts the general thrust of P1, but, unfortunately, this is not an occasion on which I can take the time to address them.

Fifthly, I should note that I am aware that there are cases of what most of us would call “self-defense,” and in some instances *legitimate* self-defense, that are not covered by P1. Examples include what is usually called the case of the “Innocent Attacker”: someone, for example, whose mind and actions are being controlled by some other party and who is used by that party to initiate an attack that that other party wants initiated against some innocent person. It may be that in such cases the innocent attacker’s intended victim has a right to defend herself, with serious harms, against the innocent attacker, if that’s

her only alternative to being harmed herself. But this, if it is so—and I think in many such cases it would be so—is not justified by P1. We need some other, quite different principle.

Here I have space to say just two things about this limitation on P1. First, it forces us, obviously, to ask whether we ultimately want, because we feel we must have, a *single central idea* in a theory of justified self-defense, or whether we can be content—perhaps because we feel we must and should be content—with a theory of self-defense that has at least two (or more) separate parts, each with a different central idea about what justifies self-defensive violence in the relevant cases. I personally think, and have argued, that we better reflect the realities of our moral thinking by appealing to two different principles for these very different sorts of cases and abandoning any search for a single, unifying principle. But I realize this is not a view that is universally shared, even by those who are sympathetic to P1 (for dealing with the cases they agree it treats correctly).

Secondly, though, note that in this context—where we are dealing with cases to which P1 does not apply—P1 does have something interesting and, I think, intuitively quite compelling to say about the specific case of the innocent attacker: it says, or rather implies, that if there is a chance to repel the innocent attacker’s attack by harming the person who is controlling his mind and actions, rather than harming the innocent attacker to repel the attack, the “controller” may rightly be harmed (and indeed is the one who *should* be harmed, if the defender chooses to defend herself), given that he is the one responsible for the fact that *someone* has to be harmed in the case as imagined (that is, given that either the innocent victim or the attacker or the controller has to be harmed, in the case as currently imagined). And this, it seems to me, as I’ve already indicated, is what most of us would intuitively be inclined to say about such cases. (I return to this point below.)

Finally, in connection with the last point above, it’s worth noting, especially since I’ll be taking P1 so seriously as a reference-point in what follows, that the general idea behind this principle is a very

powerful one, both in the depth of its appeal and in the breadth of its application. As far as depth is concerned, here I can only note that, in my experience, many people find the idea of the justice or fairness of holding people responsible for the foreseeable consequences of their voluntary, wrongful actions, at least in cases where a foreseeable choice of harms is involved, to be extremely and very deeply appealing. And as for breadth, while space prohibits a lengthy discussion here, I should note that this idea (of the justice or fairness of holding people responsible for the foreseeable consequences of their wrongful actions) is not limited to the sorts of cases with which I illustrate the application of P1 in the present paper. To take just one example, it will be helpful to sketch briefly a case that has nothing to do with self-defense, at least in the standard sense of that term, and yet seems nicely treated by P1. Suppose you and I and some friends are committed but still novice rock-climbers. We want to make an especially challenging climb—our first really difficult climb—but for various reasons we want to complete the climb at the “half-way” point—at the top of the rock formation, that is to say—and then have a small helicopter meet us there to take us down. On the day of the climb, however, the possibility of bad weather is brought to our attention, so we arrange with the helicopter company (by way of pre-payment) to have a somewhat larger and more storm-worthy helicopter meet us, just in case the bad weather materializes.

In the event, the company’s owner, who is also the pilot of the helicopter meeting us, decides to meet us in the smaller helicopter, despite the fact that we’ve paid for the larger, stronger one, so that he can save himself some money (he assumes we won’t realize what he’s done). Just as he lands to pick us up, a fierce storm arrives, but not so fierce as to keep some of us from getting down before things get worse. The problem is that, in a storm, the smaller helicopter will safely accommodate only five people, whereas there are six of us, including the pilot/owner, who need to get down. So someone will have to stay on top of the rocks, and, given the weather, it’s likely they will be there for the night. The

pilot/owner suggests that the rest of us use a lottery to decide which of us (excluding him) stays behind, and, given the weather, he suggests we do so quickly.

As it happens, however, one of our number (one of the climbers, I mean) is himself an experienced helicopter pilot; so, when the company's pilot (and owner) tells us that one of us will have to stay on the rocks, and suggests, as a decision-procedure, a lottery that does not include him, our fellow-climber points out that in fact *he* can pilot us down, with the company's owner/pilot spending the night on the rocks. And he points out, as well, that, all other things being equal, this seems like the fair solution to our problem, since it was he—the company's pilot and owner—who chose to arrive in the smaller chopper, despite our having paid for the larger, stronger one and despite his knowledge of the fact that there was a distinct possibility of bad weather, in which case (if the bad weather materialized) one of the group would have to be left on the rocks for the night.⁶

It seems to me that P1, or a very plausible extension of P1, supports the substitute-pilot's reasoning, given that P1 holds that when one person knowingly and wrongfully makes it the case that either he or some other person must be harmed, justice (or fairness) allows the latter to choose that the former is harmed, rather than that she is harmed, at least if the harms involved are roughly proportionate. (I am assuming it was wrong of the owner to use the smaller helicopter, given our pre-payment and given the imminent risk of bad weather.) To be sure, we have to count being left on the top of the rocks as a harm for this application of P1 to work, but it's easy enough to construct our case so that that is so or, alternatively, to read P1 as including risk of harm as well as near-certain harm.⁷

So, while I shall use it here solely as a principle for the justification of self-defense, it's important, given how seriously I take this principle in what follows, to note that P1 is actually a principle whose application ranges much more widely than my use of it here suggests. This fact about P1, I assume, adds to its appeal as a moral principle, and in fact adds considerably to its appeal, in my opinion, at least

intuitively. For it suggests, though of course it does not prove, that P1, far from being some ad hoc device for “justifying” self-defense, is actually a principle that appeals to moral intuitions that range much more widely than those involved in the cases to which we shall be applying it, in what follows, might seem to suggest.

So much for brief commentary on P1. Suppose, whatever you think of my most recent remarks, you find P1 tempting, *at least* as a basis for explaining many commonly-held intuitions about when harming another in *self-defense* is justified, and suppose, further, you are inclined to accept it as a plausible basis, independently of commonly-held intuitions, for the justification of harming another in self-defense in the cases to which it applies. I noted above that in fashioning P1, and the general views about self-defense that it embodies, I was thinking about fairly ordinary, uncomplicated cases of self-defense and certainly not thinking about the sorts of “contrived” cases that might interest someone thinking about the doctrine of “*Actio Libera in Causa*.” I now want to turn to a consideration of some of these latter kinds of cases and ask what, if anything, my views about how to treat standard cases tell us about how to treat these non-standard, “contrived” cases. (Recall that I am not talking about how we ought to treat such cases in the law. I’m simply asking what, if anything, a putative *moral* principle like P1 tells us about how to treat the relevant (“contrived”)) cases from the moral point of view.)

Imagine, then, someone who badly wants to hurt another (specific) person but who does not want to do so without first protecting himself, legally, from the ordinary legal consequences of harming another without just cause. He conceives a plan: he will wait for a situation in which he can provoke this person into attacking him—let us suppose he knows her well enough to know the right thing to say—and then resist her attack by harming her in just the way he would have done had she initiated an attack unprovoked. Then, if taken to court for his defensive actions, he will plead “self-defense.” (Let it be the case that what he independently wants to do to her is exactly what he will have to do to protect himself

from her anticipated attack and that the amount of harm involved would not be radically disproportionate to the harm she will do to him if her attack is not successfully repulsed.) Quite apart from how the courts may treat this individual, when he presents his planned defense in court, how does P1 suggest we think about what's going on in a case like this? That is to say, what are the implications for a case like this, *morally* speaking, if we accept a principle of moral judgment like P1?

Note, before we try to answer this question, a complication. We've imagined a case that not only involves the sorts of strategic machinations that make standard "*actio*" cases interesting but that also involves provocation. And provocation, of course, both in law and, I think, in ordinary moral thought, is a vexed issue.

Now, I want to see if I can avoid at least some of the complications introduced by the fact of provocation, for a while, at any rate, by making the following two assumptions about our case. First, let us assume that our manipulator has been careful to resort to a provocation that is quite mild, from a "reasonable person" point of view, but that he suspects will lead to an unusually strong reaction in the person he intends to provoke. Secondly, let us assume that while the manipulator can confidently expect a strong reaction to his remarks, this will not be a reaction that the person he aims to provoke will honestly be able to say she could not resist engaging in. On the contrary: for our purposes it's important to imagine, in this first version of our case, that even *she* (M's "victim" or "target," whom I shall henceforth call "T") would admit that she could have refrained from reacting as she was expected to react but that, under the circumstances, she chose not to do so. She was simply so indignant, she might say, at what the manipulator said, that she decided to teach him a lesson.

What does P1 tell us about Manipulator's self-defensive rights in such a case? What does it tell us about his moral right to defend himself in this situation, that is to say, quite apart from anything we might want to say about what his legal rights would or should be?

If we ignore for a moment the wider context of our case, which includes the fact of M's plan to get himself into a position to harm T and then plead "self-defense" at trial, and focus *just* on T's reaction to M's remarks and M's reaction to her reaction to those remarks, it seems, apart from possible complications due to the fact that provocation was involved, that P1 supports at least some degree of harmful self-defensive activity on M's part. After all, we are assuming that T *chose* to attack M, "to teach him a lesson," and that the only way M could protect himself from her attack was to harm her in some way. (I am also assuming, of course, to get this result, that T's aggressive reaction to M's remarks was a *wrongful* reaction, given that his remarks, while obnoxious, were not such as to warrant a physical assault of the kind T in fact initiates. I am stipulating, that is to say, that, from a moral point of view, T would not be justified in attacking M, in the way she does, *at least as long as we are ignoring the larger context of his remarks*, in which they are part of a plan to harm her by getting her to harm him.)

What about the fact that M has provoked T, though? Unfortunately, P1 doesn't say anything about provocation and its implications (and this, I now realize, is a serious problem with P1). We have, of course, assumed for the sake of discussion that M's provoking remarks are relatively mild—such that the stereotypical "reasonable person" would not find them *unduly* provocative—and that T's response, by her own imagined admission, was chosen and not somehow mechanically elicited from her. Still, even leaving M's larger strategic plan out of consideration, the fact of provocation might have a bearing on some individuals' thinking about this case (I mean the case "narrowly conceived," so that we are for the moment continuing to ignore the larger plan in which M's provocative remarks are embedded).

I think that for present purposes we have to set this worry (about possible ongoing concerns about the implications of provocation, despite our assumptions about the moderate nature of M's remarks and the voluntariness of T's response) aside. I say this not just because, given our assumptions, it's hard to imagine what more we could say to defuse concerns about the relevance of provocation, but also

because I don't think that whatever we say about provocation, at this point in our analysis, will make a difference to what we decide to say about the case that actually interests us: that is, the case in which we enlarge our focus and consider the possible relevance, to someone who accepts P1, of M's strategic *plan* to get himself into a position to harm T and then make a legal plea of justifiable self-defense. My thought is that, whatever the relevance of provocation in one's views about the narrowly-conceived version of our case, these views, will not, by themselves, determine what one says about the relevance of the fact that the provocation is itself part of a larger *plan*, on M's part, to get T to attack him so that he can harm her in "self-defense." But more on this below.

So, if we look at M's behavior narrowly, focusing just on his remarks and T's reaction to those remarks, and if we ignore for now any ongoing concerns about the relevance of provocation, it appears that P1 gives M the right to resist T's attack, at least within the limits set by P1's proportionality requirement. After all, from this "narrow" or limited point of view, it's T who has wrongfully made it the case that either she must be harmed or M must be harmed, and in such circumstances P1 gives M the right to resist being harmed (again, provided his resistance stays within the limits set by P1's proportionality requirement). (M's provocative remarks, we might say, for those still concerned about the possible relevance of provocation here, gave her ((we could even say *forced* on her)) an opportunity to think about "teaching M a lesson," perhaps, but they certainly did not make it the case that one of these two individuals had to be harmed, physically, in some way).⁸

Of course, this (above) will, I suspect, be, at least initially, of little interest to those whose main concern is the problem of the "*Actio Libera in Causa*," even if they are willing to ask, with me, what light, if any, a principle like P1 sheds on this problem as applied to the sort of case that currently interests us. For what *they* will want to know is not how P1 would treat our case independently of its larger context but,

rather, whether that larger context is relevant to the application of P1 once we bring it (the larger context) into play.

So what should we say in that case—our case, that is, in all its imagined complexity? A simple-minded answer would be that, once we bring the fact of M's plan into play, P1 deprives him of any right to harm T in self-defense. For, on a simple-minded view, it's M who has wrongfully (we are assuming), through his machinations, made it the case that someone must be harmed, and hence, according to P1, it's he who must bear harm rather than his innocent victim.

But, of course, this ignores the fact that, as we have constructed the case, T is not an entirely innocent or blameless victim. Indeed, making this clear was the point of our discussion above, in which we first looked at our case from what I called the "narrow" or "limited" point of view. To be sure, T is a "victim," as we might say, of M's plot to get her to attack him. But in the case as we've constructed it, she makes a free choice to harm him—in order to "teach him a lesson"—and hence must be said to bear at least some of the responsibility for the fact that either M gets harmed or she gets harmed. (Note, for the sake of clarity, that there *is* a case in which what I've called the "simple-minded view" is not simple-minded but, rather, on the mark. That's a case in which T's response to M's provocation is entirely out of her control and in which M is counting on the fact that this is so. In such a case, it *would* be true that M is ((solely)) responsible for the fact that someone must suffer a certain degree of harm—either he or an innocent victim--and hence P1 would deny him the right to defend himself against the attack against himself that he has caused. But this is not the case with which we are currently concerned.)

In the case with which we are currently concerned, one might say, we seem, at least *prima facie*, to have a curious kind of shared responsibility: we have stipulated that T is responsible for choosing to physically attack M because of his insult (I am assuming, as indicated earlier, that, taken in isolation from the fact of M's strategic plan, such an attack is not warranted, morally, even if dressing him down verbally would

be warranted), but we have also stipulated (or are assuming) that T is in a situation where she has occasion even to *consider* attacking M only because of M's decision to put her in such a situation in the first place—that is, only because of his decision to put her in that situation in order to have the opportunity to harm her and then escape the legal consequences of doing so by claiming “self-defense.” The problem, on this way of thinking about our case, is how to sort out and characterize the respective responsibilities of each party and then figure out what someone sympathetic to the fundamental ideas inherent in P1 should say about M's putative right to resist T's attack (*given that we are now considering the full context of the case, as originally imagined*).

I shall attempt to address this problem in a moment. First, though, it will be helpful to confront the following rhetorical question, which we can imagine an opponent of the “shared responsibility” view asking: “In what sense does the fact that M laid the plot he laid make him even *partly* responsible for the fact that someone must be harmed in the case as we have imagined it?” “To be sure,” this critic might say, “M is a blackguard, and perhaps there ought to be laws against the kinds of machinations in which he has engaged. But from the standpoint of the sort of responsibility P1 requires, it's hard to see how M can be said to be even *partly* responsible, morally, for the fact that T has put him in a situation where he must harm her or be harmed himself. After all, it was *her* (voluntary) reaction that put him in this situation, not *his* desire to elicit that reaction.” (Recall that we are holding aside for now concerns about the relevance of M's provocative remarks, in our case as narrowly conceived. Our interest here is in the relevance of the fact that he has *planned* those remarks as a way of eliciting a response that will require him to harm T in a certain way, if he wishes to remain unharmed himself, and then be able to argue “self-defense” when she takes him to court for thus harming her. And the question posed by the opponent of the shared responsibility view is: *why should we suppose that this kind of ((let us suppose)) wrongful strategic planning makes M responsible, in the sense required by P1, for its being the case that someone must be harmed—either he or some innocent target of his?*)⁹

The negative (anti-shared-responsibility) view suggested by this question is a tempting view, it seems to me—in fact, it’s what I earlier called the “simple-minded” view about our case, but now no longer appropriately so-called, since it has taken relevant (larger-context) considerations into account and, by its own lights, addressed them. But it is also a troubling view, inasmuch as it seems unreasonable to deny that M has any responsibility whatsoever, in the relevant sense of ‘responsibility’, for its being the case, in the situation seen as a whole, that someone, either he or his target, must be harmed. After all, if he hadn’t formed and acted on his plan, there would be no need for anyone to be harmed.

So, there are at least two possible and not altogether implausible views about the relevance of M’s plan (to get T to attack him, so that he can harm her in “self-defense”) to the question of what his self-defensive rights are, morally speaking, from the standpoint of a principle like P1: the “shared responsibility” view holds that, in some sense, both M *and* T are responsible for the fact that M has ended up in a situation where either M must suffer harm or T must suffer harm, while the critic who rejects this view holds that, whatever M’s other responsibilities and moral flaws, it is unreasonable to hold that, simply because of his (successful) plan, M must share the responsibility for the fact that he is in a situation where either he must be harmed or T must be harmed. (Recall that we are holding aside the relevance to M’s putative self-defensive rights the fact that he has provoked T, hoping to get her to attack him. Our question is: apart from the relevance ((to his self-defensive rights)) of the provocation, what is the relevance, if any, of the fact that those remarks were part of a *plan* aimed at getting her to attack him.)

Now, clearly, the person who *rejects* the shared responsibility view, at least as we have imagined his position, is rejecting, as well, the claim that, in planning as he has planned, M has lost the right to defend himself when T attacks him (in accordance with his plan). To be sure—and we’ll return to this below—this person need not reject various other claims about M’s responsibilities and moral standing

(including claims about various moral obligations M might be said to have because of his immoral machinations). The point here is simply a point about what the critic of the shared responsibility view chooses to say about M's right to defend himself against T's attack, and that point is that he (the critic) is perfectly free to say that M has every right to resist T's attack (within the limits of P1's proportionality requirement), despite the fact that he has done his best to bring about that attack.

What about the proponent of the shared responsibility view? Central to her view is the assumption that if M hadn't formed and acted on his plan, it's reasonable to suppose that T would not have attacked M and hence reasonable to suppose that M would not have had to harm T in order to prevent her from (wrongfully) harming him. Of course, T did attack M, because of M's remarks, and in freely doing so made it necessary for M to choose between being wrongfully harmed or harming his attacker. But this is precisely the proponent of the shared responsibility view's point: M is responsible for T's getting into a situation where she would think about attacking him—without this there would be no need for a choice of harms—while T is responsible for acting as we have imagined her acting in that situation, thus wrongfully putting M into the position of having to choose to harm or be harmed. This is what the proponent of the shared responsibility means by *shared* responsibility.

But, now, even supposing the shared responsibility view is getting something importantly right here—something the critic of this view is missing—where does her commitment to the idea of shared responsibility leave her standing on the question of M's putative right to resist T's attack, once he has succeeded, in accordance with his plan, in getting her to attack him? Where does it leave her standing on this question, that is to say, given the assumption that her decision on the question of M's self-defensive rights will be based on P1 or, at any rate, the kinds of considerations that inform P1?

P1 does not explicitly address this question, nor was it intended to do so. But given the idea of personal responsibility that is central to P1, and the idea of justice or fairness that is P1's moral engine, we ought

to be able to say *something* about what an appropriately extended version of P1 would say about M's rights in the case we have imagined, when we are imagining it as a case of shared responsibility for the fact that someone has to be harmed because of the wrongful conduct of one of those who can be harmed in order to spare harm to the other. The question, we might say, is *"What does justice, or fairness, allow in a case like this, when the sense of 'justice' or 'fairness' in play is the one that makes P1 seem so plausible, at least to some of us, as a way of understanding the standard cases it was formulated to help us understand?"*

The answer, it seems to me, is that, even granting M's role in planning and provoking T's attack, it's hard to see how justice or fairness, as we are conceiving them, could entirely deprive him of the right to defend himself from T's attack. No doubt, he's acted wrongfully (we are supposing) in plotting to get T to attack him so that he can harm her in self-defense. And no doubt there will be moral consequences—I'll discuss some of them below—attached to this wrongful act. But as for his right to defend himself, assuming justifiable self-defense (at least in standard cases) to rest on the sorts of considerations that drive a principle like P1, it's hard to see how the "contrived" nature of M's situation could, by the lights of P1, deprive M of the right to defend himself against T's attack. After all, it's T, as we've seen, not M, who, in the final analysis, has wrongfully made it the case that someone must suffer, and P1 suggests that this gives her victim the right to choose that she should suffer rather than he. To be sure, we are assuming that M bears some responsibility here—namely, the responsibility for getting the relevant chain of events rolling. But, as I've said, while that responsibility bears looking into, it's hard to see why it should affect our judgment, assuming we accept the basic thrust of P1 and its ideas of justice and responsibility-for-the-consequences-of-one's-wrongful-actions, that, given T's voluntary (and wrongful) decision, M has a right to defend himself from the harm that T is attempting to inflict on him.

Suppose this line of thought is right: though M is responsible, on the “shared responsibility” view, for its being the case that T has to even *consider* attacking him, and though setting her up in this way constituted a wrongful act (or set of acts), his responsibility in this regard is not enough, if one accepts a principle like P1, to deprive him *entirely* of his right to defend himself against T’s attack. Given our earlier discussion of the views of the *critic* of the shared responsibility view, this means that, regardless of whether we accept the “shared responsibility” view of our case or its negative alternative, we end up forced to say, as long as we stick to the considerations enshrined in P1, that M does indeed have a right to resist T’s attack, despite his role in planning and provoking that attack.

“So much the worse for P1,” someone might say, thinking that any principle that yields the preceding result must be unsound. I return to this point in my concluding remarks. First, though, notice that neither view—the “shared responsibility” view or its opposite—is deprived, by its commitment to P1, of the right to *criticize* M for his conduct *and*, perhaps, set conditions or restrictions on his right to defend himself (beyond those, that is, embodied in the proportionality requirement that is included in P1).

It’s easy enough to imagine the sorts of criticisms the holders of either of these views might choose to make of M, and, for the sake of brevity, I shall skip over the task of enumerating or imagining these criticisms here. More interesting, I think, for our present purposes, is the question of what sorts of *restrictions* a commitment to P1 might place on M’s (now presumed) right to resist T’s attack, beyond those entailed by P1’s proportionality requirement and the “minimum harm” requirement that is implicit in P1.¹⁰

Again, I think we could easily imagine restrictions of the relevant sort—on what exactly M may or may not do in response to T’s attack, given our various assumptions about his situation—coming from both the defender of and the objector to the shared responsibility view. Here, for the sake of brevity, and because I personally find it the more plausible view, I shall limit my remarks to what the defender of the

shared responsibility view might say about restrictions or limitations on what we are assuming is M's (moral) right to defend himself in a case like the one we have imagined.

What sorts of restrictions or limitations might one have in mind here? Think first about P1's proportionality requirement: the harm that M does to T, to protect himself from her attack, must not be radically disproportionate to the harm he would suffer if he did not defend himself against her attack. Though I haven't stressed this above, it should be obvious that a very *rough* calculation will be involved here: there's no universally or even widely accepted schedule of exactly what would count as "roughly proportionate" and "radically disproportionate" responses in cases of legitimate self-defense. What's more, in a standard case of (unprovoked) aggression, most of us would be likely to cut the defender a bit of moral slack with respect to exactly what the proportionality requirement allows and what it forbids. But, of course, we're currently concerned with a non-standard case of self-defense, in which the need for an act of self-defense has been created not just by provocation but by provocation planned in advance, to give the defender an opportunity to harm the attacker and then defend himself for doing so. In such a case, it does not seem unreasonable to suppose that, from the standpoint of someone committed to the shared responsibility view, M can rightly, and indeed *should*, be held to much higher standards, in connection with what he may and may not do in defending himself against his attacker, than those to which the defender is held in a standard case of self-defense. Specifically, it does not seem unreasonable, from the standpoint of someone who accepts both P1 and the shared responsibility view, to insist that M be very careful indeed not to do more to T than he needs to do to prevent her attack and, moreover, to keep the amount of harm he does to T, to prevent her attack, strictly within the (admittedly rough) limits of P1's proportionality limits. In neither respect, that is to say—honoring the "minimum harm" requirement and honoring the proportionality requirement—will someone who accepts P1 as a basis for legitimate self-defensive harms, and who accepts the shared responsibility

analysis of our case, be inclined to cut M any moral slack, given his role in the instigation of T's attack, in his estimates of just how much he may do to protect himself from T's attack.

What else might we say, as defenders of the "shared responsibility" view, and still accepting P1 as our basic principle for justifiable self-defense, about limits or restrictions on M's right to defend himself from T's attack? Consider the concept of "mitigation." Someone concerned to defend or otherwise try to exculpate T, morally, for her behavior, might well want to argue that the fact of M's plan to harm T by provoking her, and then defend himself by claiming "justifiable self-defense," is a fact that acts, at least to some degree, in mitigation (morally) of T's blameworthiness for her attack on M. Suppose this is right. What has that got to do with possible limits on M's right to defend himself from T's attack? Well, for one thing, it might bear on the question of whether it's reasonable, after the event, to hold M morally responsible for paying some of the costs of "repairing" T for the harms done to her in M's act of self-defense. Especially from the standpoint of the defender of the shared responsibility view, it would seem plausible to suppose it is a requirement of fairness, all other things being equal, to hold M liable for some of those costs. For, at least on the shared responsibility view, M is partly responsible for the existence of those costs.

So, for a defender of the shared responsibility view, there are at least two ways in which a commitment to a principle like P1 might be thought to "limit" or "restrict" M's right to defend himself from T's attack, even if we grant that that right is not altogether undermined by one's acceptance of the ideas and values that motivate P1. First, such a commitment, for someone who holds the shared responsibility view, would arguably entail a higher standard of scrutiny of M's self-defensive actions, relative both to the "minimum harm" requirement—not to do more harm than one has to do to repel an unjust attack—and to P1's proportionality requirement. Secondly, such a commitment—again, at least for someone who holds the shared responsibility view—would arguably entail that M is obligated to bear some part

(and possibly all) of the costs of T's recovery, from the harms inflicted on her by him in his defensive actions, thereby "limiting" or constraining his right to defend himself by adding, at least morally, to its potential "costs."

I suspect there are other ways in which the defender of the shared responsibility view would want to limit M's right to defend himself from T's attack, even while (as we are supposing) supporting the claim that that right is not entirely undermined, assuming a commitment to P1, by the fact that M is responsible for initiating the course of events that make self-defense necessary for him in the first place. Limits of time and space, however, not to mention limits of imaginative power, force me to cease further inquiry in this respect, at least for now, and turn, instead, in concluding, to the question of what any of this (above) might have to do with the *legal* or *jurisprudential* issues surrounding the problem of the "*Actio Libera in Causa*."

Two questions are salient in this regard, it seems to me: (i) What does our analysis of the self-defense case we've been discussing tell us about the differences between this sort of case and paradigm cases (of apparent relevance to the "*Actio*" doctrine) like the one mentioned at the beginning of this paper, where we imagined someone getting drunk in order to harm another, while drunk, and then claim that he was not responsible for the harm he has done because he was drunk when he did it? (ii) What does our analysis suggest, if anything, about how *the law* should treat cases like these—both cases like the self-induced drunkenness case, that is to say, and cases like the contrived self-defense case discussed above? I want to conclude with some very brief, and very tentative, thoughts about possible answers to these questions.

I'll begin with differences and note, first, that while both cases involve "targets"—that is, someone an ill-intentioned agent wants to harm and then defend himself for harming—there is an interesting and, on the view defended above, extremely important difference between these cases in this regard: the target

in the self-defense case, unlike her counterpart in the drunkenness case, is someone who, as part of the manipulator's plan, has to make a certain choice—namely, whether to respond to his abusive remarks or ignore them—and who, in making this choice, will be making a choice for which she can subsequently be held morally responsible. This, of course, was crucial to the outcome of the analysis defended above and would be crucial, I think, to any plausible analysis of what to say about a case like this, however much it disagrees with our conclusions above.

Related to this difference is what seems to me to be an equally important and, conceptually, extremely interesting second difference. In paradigm cases like the strategic drunkenness case, the manipulator is aiming at diminishing his subsequent responsibility for what he does by rendering himself, at least for time of the relevant act, literally “not responsible” and hence “not free” in that respect. In the self-defense case, by contrast, this is not so. On the contrary: the manipulator's plan leaves him free to respond or not respond to his target's response to his provocation. The point is simply to create a situation in which he can freely respond, with harm to his “attacker,” and then defend himself as imagined above.

To be sure, it's part of the plan that the manipulator get himself into a position in which he can argue that, given the target's attack, he had no choice but to defend himself with harm to her. But the sense in which he “has no choice” but to respond, in a case like this, is of course quite different from the sense in which, once drunk, the manipulator in the drunkenness case can say he did not make a free and responsible choice (at least at the time of action) to harm his target. (This difference could, I think, be used to argue that cases like our self-defense case should not in fact be considered cases of an “*actio libera in causa*,” since, at least implicitly, cases of the latter sort are, paradigmatically, at any rate, cases where, while the agent was free and responsible when setting the relevant course of events into action, he was, by prior intention, *not* free when he committed the harmful act to his target. This, of course, is

not true of him in contrived self-defense cases. [Note that “*Actio*” cases are in some dictionaries *defined* as cases of an action performed as a result of intended diminished *capacity*.] This is a line of argument that I think it would be worthwhile to pursue on another occasion.)

We should note, finally, that a third and obvious difference between the two sorts of cases is that the self-defense case involves provocation, with all the difficulties this raises for *any* analysis, whereas the paradigm cases do not. Unfortunately, I shall have to leave discussion of the importance of this difference, as well as of those just mentioned, for another time.

What about the law? What is the relevance of the differences just mentioned, and of our entire analysis, for that matter, to what the law should say about *either* paradigm cases like the contrived drunkenness case *or* cases like the contrived self-defense case discussed above?

At first blush, it might seem obvious what our analysis of the self-defense case, at any rate, requires us to say about cases of contrived self-defense. For it might seem obvious that if one holds that, morally, and independently of questions about the law, M would have a right to defend himself against T’s attack, then the law should recognize and respect this right, while at the same time recognizing and enforcing the limitations on its exercise discussed above. Before accepting this obvious-seeming conclusion, though, it will be worthwhile to ask, very briefly, what someone who accepts a principle like P1 for the analysis of standard or uncontrived self-defense cases might be inclined to say about the rules of law generally, quite apart from the issue of how the law ought to treat cases of contrived self-defense.

This is a very big issue, obviously, and I shall have to be ridiculously brief about it here. In essence, and as I have argued at length elsewhere, someone who accepts P1 as a basis for the analysis of standard self-defense cases will be likely, I believe, to think of *the criminal law*, at any rate, as a codification of a set of rules for *societal* self-defense.¹¹ Suppose just for the sake of argument that one accepts P1 and

also this “societal self-defense” view of the criminal law. What, in that case, would one be likely to say about cases of *contrived drunkenness* and whether the manipulator, as I have called him, should be allowed a plea of diminished responsibility in such cases?

Briefly, and without argument, I want to suggest that one would want to say that in such a case M should not be allowed a defense of diminished responsibility and should not be allowed that defense because society’s need (and right) to protect itself and its members from the sorts of ill-intentioned planning that M’s case involves. If this is right, then one will support laws of the requisite sort, and if they are passed and duly promulgated, M will rightly be held to the relevant standard and not allowed to defend himself as imagined.

What has this got to do with cases of contrived self-defense? As we saw above, it seems, *prima facie*, as though someone who accepts the conclusion of our moral analysis of such cases will have to say the law should allow M a right to defend his defensive actions in such cases, although strictly within the limits sketched above. But suppose the person who accepts our analysis above also accepts the societal self-defense view of the criminal law. Is there a plausible argument for the claim that, despite our conclusions above, society has as much right to deny the manipulator a right to plead self-defense in contrived self-defense cases as it has to deny the manipulator a right to plead self-defense in contrived drunkenness cases?

It seems to me there is not and that this is so for a reason that arises out of one of the differences between the two sorts of cases noted above: the fact that in contrived self-defense cases, like it or not, the manipulator, if he is successful, has managed to induce his target to attack him voluntarily. Presumably, we want to discourage such behavior (the manipulator’s, that is) just as much in the form of contrived self-defense plans as in the case of strategic drunkenness. But to do so in the form of laws denying the self-defense manipulator a right to plead self-defense on his own behalf seems to ignore

the fact that, if our argument above is right, and however badly we think of him, the attack he is resisting is wrongful aggression and, moreover, aggression that, apart from existing criminal laws, he would have a moral right to resist (within the limits suggested above).

This is not to say, of course, that the criminal law may not make it an offense, and a serious offense, to engage in behavior like that of the manipulator in contrived self-defense cases—in a “conspiracy,” that is to say, to induce someone to attack him so that he can attack her in self-defense. If the argument above is correct, however, the relevant legislation must be aimed at the planning and not at depriving the manipulator of his moral right to defend himself against voluntary, wrongful attacks against him.

This is, for me at any rate, not a welcome implication of the view of the moral basis of self-defense that I hold. But it does appear to me to be an implication of that view and does not, by itself, incline me to reject the view rather than embrace an unwelcome implication.¹²

¹ Claire Finkelstein and Leo Katz, “Contrived Defenses and Deterrent Threats: Two Facets of One Problem” (5 *Ohio State Journal of Criminal Law* 2008, p. 480.

² *Ibid.*, p. 480.

³ See especially “The Justification of Deterrent Violence,” *Ethics*, Vol. 100 (1990), 301-317. All of my work on self-defense and related issues of “forced choice” was importantly influenced by Philip Montague’s extremely important paper “Self-defense and Choosing Between Lives,” *Philosophical Studies*, Vol. 40 (1981), pp. 207-219, and by conversations with Montague on these issues.

⁴ Note that I am assuming here, in addition to the “proportionality” requirement that is explicit in the final clause, a “minimum harm” requirement which holds that, quite apart from proportionality, one must not do more than one reasonably believes one has to do to prevent wrongful harm to oneself. Note also, quite importantly, that P1 is actually a special case of a more general principle, which I shall have to ignore here, that supports protecting the innocent from wrongful harm regardless of whether the innocent party is the defender herself or some other (third) party.

⁵ The best account of “justice as fairness” I know of is in some of the early work of John Rawls. (See especially “Justice as Fairness,” *The Philosophical Review*, Vol. 57 [1958].) In these early writings, justice, or fairness, is identified with doing one’s part in a voluntary cooperative enterprise. I think some of the insight and power of this early account gets obscured in Rawls’ later, more famous work.

⁶ So far as I know, this case emerged from a conversation with Philip Montague years ago. But it may have earlier antecedents that I am unaware of.

⁷ Note that even in its direct-self-defense applications, P1 is ultimately about *risk* of harm, as opposed to absolute certainty of harm, in any case.

⁸ Note that there *is* a sense in which T might plausibly say she *was* faced with a choice of harms: she might think that either she harms M, physically, or else she suffers the harms of public humiliation by virtue of not responding to his insults. I ignore this interesting possibility, here, partly because it's not descriptive, in my view, of a straightforward case of "self-defense." (I also ignore, importantly, the claim that T must choose between harming M, for the sake of *detering* future insults, or living with a higher risk of further insults as a result of not harming him now.)

⁹ Note that the objector certainly need not be saying that M is not morally responsible for any sort of wrongdoing at all. He may well feel—and I think it would in any case be plausible to say—that M is responsible for a serious act of wrongdoing in "setting up" T, as he has done, and that he is deserving of serious moral censure for what he's done. His (the critic's) point, rather, is that whatever else he is responsible for, M is not responsible for wrongfully making it the case that someone has to be harmed—either he or his "target," T. For, as indicated, he (the critic imagined above) doesn't see how we can plausibly say that M is responsible for that's being the case. Rather, that responsibility, on his view, is T's. I return to this point below.

¹⁰ On the "minimum harm" requirement, see Note 3 above.

¹¹ Others, of course, have argued for this way of thinking of the criminal law. See, for just one example, Philip Montague, "Punishment and Societal Defense," *Criminal Justice Ethics*, Vol. 2 (1983).

¹² I am grateful to my colleagues Justin D'Arms and Don Hubin for helpful conversations on the problems discussed above.