*Legal Theory*, 5 (1999), 339–362. Printed in the United States of America


# PREEMPTING ONESELF:

## *The Right and the Duty to Forestall One's Own Wrongdoing*

Leo Katz

### I. THE PROBLEM OF PREEMPTIVE ACTION

The best test of whether an intellectual problem is genuinely worthwhile is whether there exists a Jewish joke on point. By that quite exacting standard, the problem of preemptive action is certainly an important one. To be sure, the joke that comes to mind is not especially funny, but that is not what counts. It is the existence of the joke, not its funniness, that matters.

The joke goes as follows: An old man and a young man find themselves sitting opposite each other in a train compartment somewhere in Galicia, circa 1900, heading toward its final stop on a Friday mid-afternoon. The young man leans forward to ask the old man what time it is, but the old man does not answer. The young man repeats the question—still no answer. He tries a third time, in a much louder voice, and points to his wrist. The old man still ignores him, and the young man gives up. A short while later, as the train is pulling into the station, the old man gets up, turns to the young one and asks him whether he mighty carry his rather heavy suitcases for him onto the station platform. The young man does as he is bidden, but on discovering that the old man is not the least bit hard of hearing, cannot stop himself from asking: "I did, you must concede, all that courtesy requires. Why could you not do the same when I asked you what time it was?"

"Young man," replied the old man, "had I told you what time it was we would have gotten into a conversation. Had we gotten into a conversation, you would have told me that you are new in this town—which you must be since I don't know you. Had you told me that you are new in this town—and evidently Jewish—I would have had to invite you for Shabbat. Had I invited you for Shabbat you would have met my family. Had you met my family, you would have met my beautiful daughter. Had you met my beautiful daughter, well, of course you would have fallen in love with her. . . . BUT I DON'T WANT A SON-IN-LAW WITHOUT A WATCH!"

This, in a Jewish nutshell, is the problem of preemptive action. Let me proceed now to a somewhat drier, more traditional statement of the prob-



https://doi.org/10.1017/S1352325299053045 Published online by Cambridge University Press

lem, which would, I think, go something like this: Gazing far into the future, I anticipate that somebody is going to do something he shouldn't, something which I will be entitled to prevent him from doing if he ever tries it: I anticipate that he is going to attack me, which of course I will be entitled to prevent by the use of self-defense and for which I will be entitled to have him punished as well.

Now suppose that so far he has not made any moves toward actually attacking me. In fact, he has not even formed the intention to do anything violent. But I have it in my power to predict with enormous statistical accuracy that he will do something wrong in the somewhat distant future. Am I entitled to shoot him to avert such wrongdoing? Many people have a strong intuition that I am not. And there is a great deal in both law and morality to support them. But there also is a great deal that goes against them. The fact is that there is something indisputably odd and problematic about this intuition that somehow we are barred from taking preemptive action against an attack we know perfectly well is coming, just because it happens not yet to be "officially" sufficiently underway. It is the oddness of that intuition which makes preemptive action not merely a practical but an intellectual problem—the problem of preemptive action.

Although simple cases of self-defense raise the problem most simply, it is in fact ubiquitous. In *Anarchy, State and Utopia,*[1] Robert Nozick offers a memorable depiction of how the problem arises in the context of international relations:

> According to usual doctrine, under some circumstances a country X may launch a preemptive attack, or a preventive war, upon another country Y; for example, if Y is itself about to launch an immediate attack upon X, or if Y has announced that it will do so upon reaching a certain level of military readiness, which it expects to do some time soon. Yet it is not accepted doctrine that one nation X may launch a war against another nation Y because Y is getting stronger, and (such is the behavior of nations) might well attack X when it gets stronger still. Self-defense plausibly covers the first sort of situation but not the second. Why?
>
> It might be thought that the difference is merely a matter of greater or lesser probability. When a nation is about to launch an attack, or has announced that it will when and if it reaches a certain level of readiness, the probability is very high that it will attack. Whereas the probability is not as great that any nation getting stronger will attack when it attains greater strength. But the distinction between the cases does not depend upon such probability considerations. For however low the probability, estimated by the "experts" of neutral countries, of Y's launching an attack on X (in the second case) within the next ten years (0.5, 0.2, 0.05), we can imagine alternatively that Y now is about to wield a superdevice fresh out of its scientific laboratories that, with *that* probability, will conquer X; while with one minus that probability, it will do nothing. . . . The device is set to be wielded within one

1. Robert Nozick, ANARCHY, STATE AND UTOPIA (1974).

https://doi.org/10.1017/S1352325299053045 Published online by Cambridge University Press

> week; Y is committed to use it, the timetable is being followed and a countdown has begun. Here X, in self-defense, may attack, or issue an ultimatum that if the device is not dismantled within two days it will attack, and so on. . . . If Y were spinning a roulette wheel and with probability 0.025 the damage of war would be inflicted on X, X could act in self-defense. But, in the second case, even when the probability is equal, X may not so act against Y's arming. Therefore, the issue is not merely a matter of how high the probability is.[2]

The problem of preemptive action arises as well when society has to decide whether it is all right to detain dangerous individuals preventively if all we have is statistical evidence of their dangerousness. Or when society has to decide whether to extend a prisoner's sentence for crimes he has already committed in light of the crimes he is likely to commit in the future. Or when it has to decide whether to deny a suspect bail on that ground. Or, most importantly perhaps, when it has to decide whether to criminalize certain conduct that, statistically speaking, is likely to lead to crime down the road, even if in and of itself it is unobjectionable: Our drug laws in good part rest on the supposition that allowing free drug use is going to precipitate crime some time in the future in a variety of ways that aren't hard to imagine. We punish drug use, it seems, not because it is wicked but because it might *lead* to wicked things. We punish it preemptively.

I will, for purposes of this essay, take it for granted that acting on statistical predictions in the above kinds of circumstances is impermissible. I will take it for granted that at least for a non-consequentialist, that is, a deontologist (or libertarian, or retributivist—I pretty much use those terms interchangeably for now), the person who acts on reliable statistical data is doing something akin to engaging in those impermissible utilitarian trade-offs, the repudiation of which is the hallmark of being a deontologist: One cannot do various things to an innocent person to achieve various good ends thereby, including the prevention of identical sorts of harm to other innocents; one may not carve up one person to use his organs to save many more innocents from dying; one may not execute one innocent to deter crime against many more innocents; one may not torture an innocent child to get his terrorist father to reveal where he hid his bomb.[3]

To be sure, acting on statistical data is not the exact same thing as carving up an innocent to save many more innocents, because it involves proceeding against someone who is only temporarily innocent and who according to reliable statistical data that we have is going to turn vicious sometime down the road. But he is innocent for now, and for that reason it appears we are barred from proceeding against him. The principle at work here seems akin to the one that underlies the criminal law doctrine of attempt. According to the law of attempt, we are not allowed to punish a criminal

2. *Id.* at 126.
3. *See* Thomas Nagel, THE VIEW FROM NOWHERE 175–80 (1986).

https://doi.org/10.1017/S1352325299053045 Published online by Cambridge University Press

unless he has crossed a certain hard-to-define line between "mere preparation" and the actual attempt of a crime. The non-consequentialist rationale for that distinction is that up until the moment that the defendant has crossed this line he is still somehow "within his rights"; and regardless of how likely he is to transgress beyond that line we cannot touch him until he does.

Let me reiterate, however: My aim in this essay will not be to investigate the soundness of the intuition that prohibits the taking of preemptive actions against potential wrongdoers. Instead, I will take up a closely related problem, which I believe it is worth taking up for several reasons. First, because it sufficiently resembles the "classic" problem of preemptive action, and it thus seems only natural to consider it at a conference devoted to a broader consideration of this topic. Second, because it is intrinsically interesting. And third, and perhaps most importantly, because it turns out that an understanding of this related problem can shed unexpected light on the "classic" problem.

The best way to state the related problem I want to take up is with an example, which I encountered in William Manchester's biography of General Douglas MacArthur. MacArthur loved to tell a story about his father, Arthur MacArthur, also a soldier, and a bribe he almost took. The elder MacArthur was then an army captain stationed in New Orleans. A cotton broker approached him, desperately hoping to secure the temporary but illegal use of some army transport facilities. "The bribe was to be a large sum of cash, which was left on [Arthur's] desk, and a night with an exquisite Southern girl. Wiring Washington the details, Arthur concluded: 'I am depositing the money with the Treasury of the United States and request immediate relief from this command. They are getting close to my price.' "[4]

Incidents like this have attracted the attention of philosophers writing about weakness of will and of economists interested in rational choice.[5] They also deserve the attention of moral theorists, although they have gotten much less from those quarters. To see the interesting moral questions lurking in this example, let us imagine that Arthur MacArthur's superiors refuse his request for a transfer, and let us suppose further that he is dead serious in his fear of yielding to temptation. Suppose that in his desperate concern to forestall his own future wrongdoing, he deliberately commits some minor infraction that causes the army to punish him by transferring him to another, as it happens, less tempting post. Given his reasons for committing the infraction, has he acted wrongly? This is the question I shall take up in the next section, appropriately titled "The Right to Forestall One's Own Wrongdoing."

4. William Manchester, AMERICAN CAESAR: DOUGLAS MACARTHUR, 1880–1964 38 (1978).

5. *See, e.g.,* Jon Elster, ULYSSES AND THE SIRENS: STUDIES IN RATIONALITY AND IRRATIONALITY (1979); Robert H. Frank, PASSIONS WITHIN REASON: THE STRATEGIC ROLE OF EMOTIONS (1988); Derek Parfit, REASONS AND PERSONS 12–13 (1987); Thomas C. Schelling, CHOICE AND CONSEQUENCE: PERSPECTIVES OF AN ERRANT ECONOMIST 57–82 (1984).

There is a further issue lurking in the Arthur MacArthur example. Suppose that Arthur MacArthur had decided to stay at his post—that is, not contrived to have himself transferred—despite his desperate and well-founded fear about his own ability to withstand temptation. To be sure, he keeps imploring and warning his superiors to move him, but he has no success. Suppose that eventually things come to pass just as he feared they would. He is tempted and he yields. Can he still be blamed, given how hard he tried to avert this contingency? That is the question I shall take up in Section III, titled "The Duty to Forestall One's Own Wrongdoing."

Although the bulk of what follows is concerned with the morality of the actor who tries to forestall his *own* future wrongdoing—what we might call the problem of self-preemption—I will pay considerable attention to the more "classic" problem of preemption as well. I hope to convince you that thinking about the one will pay rich dividends in thinking about the other.

## II. THE RIGHT TO FORESTALL ONE'S OWN WRONGDOING

### A. Some Further Examples

My first question about Arthur MacArthur has already received some attention in the philosophical literature—not the very question, but the issue underlying it. In an article titled "Oughts, Options and Actualism," Frank Jackson and Robert Pargetter pose this hypothetical question:

> Professor Procrastinate receives an invitation to review a book. He is the best person to do the review, has the time, and so on. The best thing that can happen is that he says yes, and then writes the review when the book arrives. However, suppose it is further the case that were Procrastinate to say yes, he would not in fact get around to writing the review. Not because of incapacity or outside interference or anything like that, but because he would keep on putting the task off. (This has been known to happen.) Thus, although the best that can happen is for Procrastinate to say yes and then write, and he *can* do exactly this, what *would* in fact happen were he to say yes is that he would not write the review. Moreover, we may suppose, this latter is the worst that can happen. It would lead to the book not being reviewed at all, or at least to a review being seriously delayed.
>
> Should Procrastinate accept the invitation to review the book? Or if we suppose that he in fact declines—perhaps because he knows that he would not get around to writing the review—did he do the right thing in declining?[6]

In responding to a hypothetical question like this, philosophers tend to split into two camps, a group commonly referred to as the "actualists," who should more appropriately be called the "realists" (as in *realpolitik*) and a

6. Frank Jackson & Robert Pargetter, *Ought, Options, and Actualism,* 95 PHIL. REV. 233–55 (1986), at 235. See also the superb discussion of this topic in Michael Stocker, PLURAL AND CONFLICTING VALUES 96–109 (1990).

group called the "possibilists," who should more appropriately be called the "idealists" (as in "starry-eyed idealist"). (In fact, from hereon out I will call them what I believe they should be called.) The idealists say that "the fact that Procrastinate would not write the review were he to say yes is irrelevant. What matters is simply what is possible for Procrastinate. He can say yes and then write; that is the best; that requires *inter alia* that he say yes; therefore, he ought to say yes."[7] According to the realists, however, "the fact that Procrastinate would not actually write the review were he to say yes is crucial. It means that to say yes would be in fact to realize the worst. Therefore Procrastinate ought to say no."[8]

Neither the Professor Procrastinate hypothetical nor my Arthur MacArthur example may be weighty enough to trigger all the right intuitions. Let me suggest a more dramatic alternative to make the same point. Imagine a firefighter, let's call him Leo, who is asked to enter a burning building where two of his comrades are trapped. Others have shouldered their share of risk during this mission. It rightfully is now Leo's turn. Leo anticipates that if he tried to rescue them, his cowardice would make him fail. He foresees that when he is at the spot where the two are trapped, he will fail to do what needs to be done out of raw physical fear. Should Leo under the circumstances decline to go so that someone else, someone who already has risked his life a disproportionate number of times, go in and try to rescue them? The realists would say Leo should decline; the idealists would say he should not. Who is right?

Finally, let me round out my set of examples with one that will start to suggest one of the several ways in which the problem of self-preemption hooks up with the more "classic" problem of preemptive action. Consider a society besieged by rising crime rates. As crime rates go up, so does police brutality, vigilantism, and most alarming of all, the number of illiberal, draconian, immoral-seeming laws seeking to cope with the rising tide of crime. Astute observers looking ahead just a few years see some frightening possibilities. Unless crime is curbed quickly, they foresee the disappearance of bail, of the ban on preventive detention, of the probable-cause requirement for searches and seizures, and an increase in "preemptive" laws meting out punishment for conduct that is in and of itself inoffensive, like drug use, loitering, or the possession of potentially dangerous weapons. In other words, they foresee a plethora of laws that simply disregard our moral aversion to taking "preemptive action."

Lawmakers who dread the possibility of such draconian laws consider forestalling them by taking somewhat more draconian action now than many people believe is called for: They would like to pass somewhat draconian laws now—our current drug laws again come to mind—to forestall having to pass even more immoral, draconian laws in the future. They are

7. Jackson & Pargetter, *supra* note 6, at 235.
8. *Id.*

trying to preempt their own future wrongdoing. What about the morality of this kind of self-preemption? We know that "realists" would applaud it, and "idealists" would repudiate it. Who is right?

## B. The Gateway Sin Paradox

I believe we can solve this problem by considering something I call the "gateway sin paradox." It is a paradox only in the sense that it is counterintuitive. But there is nothing illogical about it. As far as I can tell, it is perfectly true.

The paradox involves a comparison of the blameworthiness of three characters: Saint, Bystander, and Revolutionary. Early on in life, Saint has committed an important sin, his original sin, we might say. Later on, he performed many great and heroic feats. Bystander has done neither: He has neither sinned, nor performed any great and heroic feats. When their lives are over, who is in the better moral position? Whose is the better moral ledger? The answer is that at this point we cannot tell. It all depends on how bad Saint's original sin was and how great and heroic his great and heroic feats were. But it certainly is possible—indeed, it is easy to imagine—that the sin was not so bad and the great and heroic acts were in fact so great that when all is said and done Saint ends up with a better overall moral score than Bystander.

Consider next Revolutionary. Early on in life, Revolutionary was confronted with a most peculiar choice. He was told—by whom and why I will leave unspecified—that he had to choose between two life-paths. He could either agree to commit the same original sin that Saint committed, and would then have the opportunity to perform many great and heroic feats of just the kind that Saint got to perform. Or he could decline to commit Saint's original sin, but would then never have the opportunity to commit those feats. Suppose Revolutionary decides to commit the "gateway sin" and follows it up with all of Saint's great and heroic feats. How does he compare with Saint and with Bystander? It would seem that he is no worse than Saint and maybe a little better. He committed Saint's sin and Saint's feats, but unlike Saint he committed the sin *in order* that he might perform those feats. Arguably that makes the sin a bit less sinful. And because we assumed Saint to be better (overall) than Bystander, we now know that Revolutionary is better than Bystander as well.

Things turn paradoxical if we next ask the following question: *Should Revolutionary have committed the gateway sin?* Let us make the situation just a bit more concrete. Let us assume the sin in question to be the torture of a child. Let us also assume the great and heroic feats to consist in the saving of several lives (ten altogether) under truly frightful circumstances, posing much risk, calling for great effort, in short, involving the kind of sacrifice only saints are prepared to make—enough of a sacrifice so that we can say

https://doi.org/10.1017/S1352325299053045 Published online by Cambridge University Press

that the Saint who both tortured a child early in his career and saved ten lives later in his career still comes out ahead of the Bystander who did neither. Does this mean that the Revolutionary should be willing to torture a child if that is what is necessary to enable him to heroically save ten lives? The answer for a deontologist, of course, is no. What we have here is the quintessential forbidden trade-off. Torturing a child is not permissible to save ten lives. (If you believe this is not a good example of a quintessential forbidden trade-off, you should have no trouble substituting your own and adapting the analysis accordingly.) Therein lies the gateway sin paradox: Revolutionary *should not* commit the gateway sin. But he will end up with a better moral ledger if he does.

It is worth pausing here and noting just how strange a conclusion this is. Imagine Revolutionary and Bystander both trying to decide whether to commit the gateway sin. Imagine that God appears to them at that moment, willing to dispense advice. He would of course tell them that they are not allowed to commit the gateway sin, that saving ten lives in the future does not make it all right to torture a child now. Suppose next that Revolutionary flouts God's specific command and Bystander heeds it. When the two of them finally knock on the pearly gates, God will nevertheless end up ranking Revolutionary ahead of Bystander and might even grant him admission to heaven while denying it to Bystander.

Let us try to understand what gives rise to this bizarre state of affairs. It arises because of the peculiarly asymmetrical role that aggregation plays when a deontologist decides what to do and the role it plays when he or she evaluates someone's overall moral record. When a deontologist makes a decision, aggregation is often impermissible: Good consequences cannot outweigh certain bad means. That is why some trade-offs are tabu. Nevertheless, even a deontologist cannot avoid engaging in some kind of aggregation when assessing someone's overall blameworthiness—for instance, to decide how much retribution someone merits who has committed multiple crimes. Thus, when deciding whether the Revolutionary should commit his gateway sin, the deontologist would not approve of aggregation. But when comparing the Revolutionary's overall moral record with that of the Bystander, the deontologist cannot avoid it. Hence the paradox.

Equipped with the lessons of the gateway sin paradox, let us revisit our various examples. Note how similar Arthur MacArthur's position is to that of Revolutionary. If he commits whatever minor wrong it would take to get himself transferred to a different, less temptation-laden post, he is committing something very much in the nature of a gateway sin. It will ensure more virtuous behavior down the road than if he were not to engage in the gateway sin. It thus quite clearly will improve his eventual moral ledger. But does it follow that he should commit the minor wrong? Based on the gateway sin analysis, we can conclude—probably not. Our ambivalence, our uncertainty about his position derives from our prior inability to reconcile two contradictory impulses. The perfectly sensible impulse is that he is

engaged in wrongdoing which is not outweighed or eradicated or justified by the wrong averted down the road—the same impulse, that is, which leads us to say that it is not all right to torture a child to avert several more child-torturings down the road: or to carve someone up for his organs, to avert several more such organ thefts in the future; or to execute an innocent now so as to forestall the execution of more innocents later on. The contrary impulse we feel is rooted in our sense that somehow MacArthur will end up morally ahead if he gets himself transferred. And that sense stems from the justifiable assessment that MacArthur's moral ledger will eventually look better if he now does the wrong thing and thus averts doing the wrong thing. Both impulses make sense. And the gateway sin paradox allows us to reconcile them.

The analysis of Professor Procrastinate proceeds in the same way. The "idealists" are right to say that Professor Procrastinate should not turn down the book-reviewing assignment. They are correct in refusing to accept his fear of future dereliction as grounds for turning down the assignment. To do so, we now know, is the equivalent of doing a wrong now so as to avert committing several more or more serious such wrongs in the future. The realists in turn are right in the sense that Professor Procrastinate will end up with a better moral ledger—with a better chance of entering heaven!—if he turns down the assignment.

Finally, there is the assessment of lawmakers thinking about adopting draconian legislation now—like our current drug laws—to avert more draconian laws in the future. They too are thinking about a gateway sin. They should not commit it, but if they do they will end up morally ahead. Note how ambiguous the position is in which a deontologist now ends up vis-à-vis preemptive actions. He has to condemn them, in a sense, but he is also able to approve of them in another. While disapproving of an individual instance of preemptive action, the deontologist cannot help conceding that a society that engages in a given instance of preemptive action might well end up morally ahead of a society that does not do so!

## III. THE DUTY TO FORESTALL ONE'S OWN WRONGDOING

Let us turn then to the second of my two questions about the Arthur MacArthur example. My first question—the one we just got through answering—was whether MacArthur had a right to anticipate and preempt his own wrongdoing, his weakness of will in the face of temptation, with some preemptive wrongdoing. My second question was what sort of credit he gets if he does not take such preemptive action, or rather, if all he does is his duty; he stays at his post, and as a result encounters the temptation he so hoped to avoid—and yields to it. That turns out to be a surprisingly difficult question. I should admit right at the outset that I won't actually succeed in resolving it. I will, I think, shed light on it. I will, I think, make you

appreciate the difficulty of it more fully. I will, I think, show you what implications the answer to this question holds for the more "classic" version of the problem of preemptive action—that is, the preemption of other people's wrongdoing. I just cannot answer it.

To gain a full appreciation of what makes it such a difficult issue, let me lay out a more elaborate hypothetical than the one about Arthur MacArthur. It is basically a highly stylized, modernized version of the famous episode about Ulysses and the sirens: Ulysses, you will remember, knew that sailors were apt to be seduced by the song of the sirens to crash their ships into the sirens' cliff. He therefore ordered his men to stop their ears with wax. But because he wanted to hear the sirens' song himself, he had himself tied to a mast, and he instructed the crew to ignore his orders until they had passed the cliff.

My version of this story involves an experimental neuroscientist whom I would have liked to call Dr. Ulysses, but because that is too unwieldy, I will just call him Dr. U., or better yet, Dr. Yu. Dr. Yu, who is both an experimental neuropsychologist and a biochemist, knows of a certain herb that he believes has the potential of making people quite violent. It also has the potential, if suitably refined, of being a life-saving drug of enormous significance. The key is to find a way to "detoxify" it, so as to eliminate its violence-inducing properties. Dr. Yu is rightly convinced—let us suppose—that he is most likely to come up with an idea for detoxification if he can get to experience the violence-inducing properties of the drug firsthand. What's more, he thinks it important to experience them in a social setting, where they are most likely to manifest themselves. He therefore hires several very reliable bodyguards, takes a dose of the drug, and attends a social event. The idea, of course, is that as soon as he shows signs of becoming dangerous to bystanders, his bodyguards will seize him and render him harmless.

This is in fact roughly how it plays out, except that Dr. Yu gets a bit closer to doing harm than anyone reckoned. Within minutes of the drug's full effect, he develops an intense anger toward people around him, and when an obnoxious child starts to scream in his near-vicinity, he manages to get hold of a sharp object and to hurl it in the direction of the screamer, with the words "I am going to kill you, brat!" Luckily his bodyguards succeed (just barely!) in deflecting the missile.

When later he describes his state of mind during this outburst, he confesses that he really did want to kill the child. He also reports that he did not feel the least bit insane. He felt, he says, the way he imagines many ordinary, short-tempered killers feel just before they do something terrible. The clues, by the way, which he gleans from this experiment prove invaluable in coming up with a detoxified version of the herb.

The question I want to raise is whether Dr. Yu is guilty of attempted murder. As a matter of criminal law doctrine, the answer is a simple Yes. He acted with an intent to kill. He went beyond mere preparation. He was sane

enough. Is that the answer, however, which the law *should* give? Does it track morality? Is Dr. Yu blameworthy under the circumstances?

The case AGAINST blaming him is formidable. Here are the chief arguments as I see them:

*First.* Blaming or punishing the likes of Dr. Yu would seem to defeat the point of blame and punishment. Dr. Yu embarked on a harmless course of action, which had the prospect of generating significant benefits in the long run. Isn't that just the sort of conduct we want to encourage? It is true that his course of action included conduct that, considered in isolation, would count as immoral. But the conduct did not occur in isolation. Lots of perfectly moral behavior contains subsidiary pieces of behavior that, if considered in isolation, would be considered immoral—a killing in self-defense, for instance; or a bank robbery that is abandoned midway through—but we think that the total picture is what is relevant.

*Second.* Even as a matter of criminal law doctrine, it takes but a slight and obviously necessary emendation of existing doctrines to get Dr. Yu off the hook.

As it stands, the criminal law acquits defendants who, having attempted but not completed a crime, voluntarily change course, renounce their criminal purpose, and "abandon" the crime. In other words: If I break into a bank, open the safe, am about to take out the money and am at that point arrested by waiting police, I would be guilty of attempted bank robbery. But if moments before the police appear, I experience pangs of conscience, and decide I am going to leave without the money, I go scot-free—because I voluntarily abandoned my criminal attempt.

To be sure, the abandonment defense is rife with obvious evidentiary difficulties, as well as conceptual ones: Just consider the notorious casebook example of the man who attempts a rape but "abandons" it when he realizes that the victim is wearing maternity clothing. Should his abandonment of this now "unattractive" target count as voluntary for purposes of the abandonment defense? And what about the assassin who fires two bullets at his victim, misses, but *then* has a change of heart and decides to let him live? Does this assassin too get the abandonment defense? Notwithstanding these difficult borderline cases, however, the abandonment doctrine jibes pretty well with our moral intuitions.

Dr. Yu's case is not one of abandonment, for it was third parties and not Yu's change of heart that saved the life of his victims. But of course it was Yu's prior measures that insured that such third parties would be on hand. In a sense, he abandoned his attempt before it ever started. Such cases of preemptive, "pre-crime" abandonment are not covered by the current doctrine. But it would seem like a small and salutary modification to expand the doctrine to include them.

*Third.* Consider the way the criminal law generally treats cases in which someone commits a crime while in some reponsibility-impaired state, like the driver who has an epileptic seizure and runs over a pedestrian, or the

sleepwalker who murders her roommate using an ax, or the faith healer who during one of her trances plunges a knife into her patient. The general approach to such cases is to not hold the defendants liable for anything they did while they were in their responsibility-impaired state, but to inquire carefully whether they had any inkling of what was coming, and if they did whether they took adequate precautions to forestall it. In other words, did they know that they were prone to epileptic seizures, sleepwalking or violent acts while in trance? If so, they were probably reckless in going for a drive, in not sleeping with their doors locked, and in not doing their faith healing with someone to watch over them. We look not at the moment at which someone committed his or her misdeed but at the time preceding it. That suggests we do the same thing here: that we treat what Dr. Yu did in the same way we treat an epileptic seizure, a sleepwalking episode, or the faith healing incident and that we focus on the conduct preceding and precipitating it and decide whether that conduct was culpable. Dr. Yu's conduct leading up to his violent outburst was not the least bit culpable. His ingestion of the herb that precipitated the outburst was completely unobjectionable given the safety precautions he was taking and given his high-minded reasons for ingesting the herb. Thus, he should be acquitted.

*Fourth.* If we are going to blame Dr. Yu, it seems consistency might then require us to blame some defendants who seem truly beyond reproach. Consider the soldier who likes to energize himself by imagining the enemy he is firing at to be a superior officer he particularly despises. Let us assume that for some microseconds he in fact manages to sustain that illusion. For that matter, imagine it to be the case that lots of people when they perform violent acts for which they have good and fully justifying reasons—surgeons, even, when they cut into someone's body—experience recurrent intervals during which they do not contemplate or remember the good and justifying reasons they have for doing what they are doing. In other words, during those intervals, their state of mind is really indistinguishible from that of a criminal doing the same thing for bad reasons. If Dr. Yu can be blamed, must not they be too? And that seems absurd.

Alas, there is a formidable case on the opposite side as well, which I will present by responding to each of the above arguments one-by-one.

*First.* It is tempting but wrong to think that if a course of conduct is morally justifiable, each of its segments is morally justifiable. It is thus tempting but wrong to believe that just because Dr. Yu embarked on a morally justifiable course of conduct when he decided to swallow the violence-inducing herb, what he did while under the influence of that herb is also morally justifiable. To be sure, sometimes the larger context does wash away the sinfulness of seemingly sinful conduct: Context can turn a wrongful killing into rightful self-defense. Context can turn an attempted crime into an abandoned folly. But this is not guaranteed to happen just because the defendant was morally justified in the course on which he embarked.

To see this clearly, let us switch back for the time being to the parallel

problem of rational action. Consider a chess player who is committing what looks like a colossal blunder, the ostensibly foolish sacrifice of a major piece. He is acting with seeming irrationality. Suppose it turns out that what he did in fact leads him to swift victory and that he planned it that way. His seemingly irrational actions become rational once we recognize them to be a component of a rational overall strategy.

But not all irrational actions become rational once we recognize them to be a component of a rational overall strategy. A very clear example of actions that do not is to be found in an example jointly concocted by Thomas Schelling and Derek Parfit (i.e., Parfit adapted it from Schelling):

> A man breaks into my house. He hears me calling the police. But, since the nearest town is far away, the police cannot arrive in less than fifteen minutes. The man orders me to open the safe in which I hoard my gold. He threatens that, unless he gets the gold in the next five minutes, he will start shooting my children, one by one.
>
> What is it rational for me to do? I need the answer fast. I realize that it would not be rational to give this man the gold. The man knows that, if he simply takes the gold, either I or my children could tell the police the make and number of the car in which he drives away. So there is a great risk that, if he gets the gold, he will kill me and my children before he drives away.
>
> Since it would be irrational to give this man the gold, should I ignore his threat? This would also be irrational. There is a great risk that he will kill one of my children, to make me believe his threat that, unless he gets the gold, he will kill my other children.
>
> What should I do? It is very likely that, whether or not I give this man the gold, he will kill us all. I am in a desperate position. Fortunately, I remember reading Schelling's *The Strategy of Conflict*. I also have a special drug conveniently at hand. This drug causes one to be, for a brief period, very irrational. I reach for the bottle and drink a mouthful before the man can stop me. Within a few seconds, it becomes apparent that I am crazy. Reeling about the room, I say to the man: "Go ahead. I love my children. So please kill them." The man tries to get the gold by torturing me. I cry out: "This is agony. So please go on."
>
> Given the state I am in, the man is now powerless. He can do nothing that will induce me to open the safe. Threats and torture cannot force concessions from someone who is irrational. The man can only flee, hoping to escape the police. And, since I am in this state, the man is less likely to believe that I would record the number on his car. He therefore has less reason to kill me.
>
> While I am in this state, I shall act in ways that are very irrational. There is a risk that, before the police arrive, I may harm myself or my children. But since I have no gun, the risk is small. And making myself irrational is the best way to reduce the great risk that this man will kill us all.[9]

In this example, someone is behaving somewhat like the ostensibly blundering chess player. But his strategic blunders do not become rational merely

9. Parfit, *supra* note 5, at 12.

because they are strategic, whereas in the case of the chess player the context washes away, as it were, all taint of irrationality; in the case of the robbery victim, that does not happen.

The above analogies would seem to carry over to morality. The person who is acting in self-defense or who abandons a criminal attempt is like the blundering chess player. His seemingly immoral actions cease to be that when placed "in context." The context washes away the sin. But there are moral analogues to the robbery victim where context does not wash away the sin, even though it seems it should. A very clear example is again furnished by Parfit:

> Suppose that I have some public career [let's say, as a prosecutor] that would be wrecked if I was involved in a scandal. I have an enemy, a criminal whom I exposed. This enemy, now released, wants revenge. Rather than simply injuring me, he decides to force me to corrupt myself, knowing that I shall think this worse than most injuries. He threatens that either he or some member of his gang will kill all my children, unless I act in some obscene way, that he will film. If he later sent this film to some journalist, my career would be wrecked. He will thus be able later, by threatening to wreck my career, to cause me to choose to act wrongly. He will cause me to choose to help him commit various minor crimes. Though I am morally as good as most people, I am not a saint. I would not act very wrongly merely to save my career; but I would help my enemy to commit minor crimes. I would here be acting wrongly even given the fact that, if I refuse to help my enemy, my career would be wrecked. We can next suppose that, since I know my enemy well, I have good reason to believe both that, if I refuse to let him make his film, my children will be killed and to believe that, if I do not refuse, they will not be killed.[10]

Here we have the case of someone who morally causes himself to behave immorally (just as the robbery victim rationally caused himself to behave irrationally). The immoral actions he takes—the minor crimes which his nemesis blackmails him into committing—are part of a course of conduct that he is morally justified in embarking on. But that does not wash away their immorality.

The same would seem to be true of Dr. Yu. We certainly cannot confidently conclude that merely because he is morally justified in swallowing the herb that he knows will cause him to behave immorally, he therefore is morally justified in behaving as he does while under the herb's influence.

*Second.* It is tempting but wrong to think that a slight emendation of the abandonment defense can get Dr. Yu off the hook. The gist of the abandonment defense is the idea of repentance. Criminals who repent improve their moral position. If they do so once their crime is complete, it is a ground for cutting their sentence. If they do so before their crime is complete, it is a

10. *Id.* at 38.

ground for not punishing them at all. But timing is of the essence when it comes to repentance. Repenting now and acting later will not work. Timing often is crucial in morality. If I intend to run someone over just before I do, it is murder; if I first run him over and only then form the intent, it is nothing. It would thus be neither a modest nor an obviously salutary reform to expand the abandonment defense to encompass cases where the abandonment precedes the attempt.

*Third.* It is tempting but wrong to draw an analogy between Dr. Yu and a driver who runs someone over during an epileptic seizure, or a sleepwalker who murders someone with an ax, or a faith healer who knifes someone during a trance, but the analogy is misplaced. If we want to charge any of those killers with, let us say, manslaughter, it is only natural to ask whether we can find any moment in time when they recklessly caused the death that eventuated. In answer to that question we are then able to say that while the epileptic did not behave recklessly when he ran someone over, he was reckless in stepping into the car; that while the sleepwalker was not reckless when she murdered her roommate, she was reckless in not locking herself into her room when asleep; that while the faith healer was not reckless when she knifed her patient, she was reckless in not getting someone to watch over her while she ministered to her patient. That sort of approach only works because the defendant was not taking any voluntary actions while she was in her special state.

That is not true of Dr. Yu. Dr. Yu is not in an epileptic state or, indeed, in any kind of responsibility-impaired state. He is acting quite voluntarily, and there is nothing wrong with our holding him responsible for such responsible actions as he takes while he is under the influence of his herb. One might, of course, ask why we focus on this moment and not some earlier one. But if we did, that would be most unlike what we do in the criminal law ordinarily.

Consider the case of a driver who steps into a car while intoxicated and in a super-aggressive mood. One of the things that would lead us to describe him as reckless is that he is now liable to do some violent intentional acts—which in fact he proceeds to do. He deliberately runs over a pedestrian. Suppose he later wants to argue that he should only be found guilty of manslaughter because we should focus on the recklessness with which he started to drive rather than the intention with which he ran over his victim. But why should we do that? Why isn't the intention with which he killed the pedestrian the right measure of his culpability? The general approach is to start with the harm that eventuated and find the nearest moment in time at which the defendant can be said to have taken a culpable action that eventuated in the bad result. That moment will then determine his guilt.

*Fourth.* It is tempting but wrong to acquit Dr. Yu on the grounds that if we convict him we will be led to acquit as well the soldier who likes to energize himself by imagining the enemy he is firing at is his sergeant, or the surgeon who imagines he is assaulting his patient. It would be wrong to be too moved

by those cases because there is an equally compelling set of cases on the other side of the issue.

If we acquit Dr. Yu, we will find it difficult not to acquit most run-of-the-mill criminals, because on close inspection it is very hard to make out a principled difference between his case and theirs. The different "feel" of the two types of cases turns out to be superficial. To see this, let me construct a series of hypotheticals that will build a bridge between Dr. Yu and the run-of-the-mill criminal.

The first plank of the bridge would be the following kind of case: Dr. Yu Jr. is born with the kind of temperament old Dr. Yu acquired by taking his drug. Yu Jr. also hires bodyguards to watch over him. Yu Jr.'s bodyguards occasionally fail him, as they do old Yu. On one of those occasions, Yu Jr. nearly kills someone. If we acquit Dr. Yu, it is hard to see why we wouldn't do the same for Yu Jr.

The second plank of the bridge is the following kind of case: Dr. Yu III does not have the resources to hire bodyguards. He does other things that are within his means: He avoids situations that might provoke his temperament. He submits to psychotherapy, frequent electroshocks, and even brain surgery—all of which helps, but not quite enough. On at least one occasion, Yu III very nearly kills someone. If we acquit Yu and Yu Jr., it is hard to see why we would not do the same for Yu III.

Now for the third and final plank of the bridge: Dr. Yu IV is just like Yu III, but he does kill someone. The bridge is now complete. Yu IV is recognizably just a run-of-the-mill wrongdoer who occasionally tries, but frequently fails, to keep his inner demons in check. If we acquit Yu, Yu Jr., and Yu III, it is hard to see why we would not do the same for Yu IV.

In building this bridge, I am not merely playing a Sorites game: I am not just taking you down a slide on the slippery slope. It is not the case that there are incremental, imperceptible differences between the original case and those that form the planks of my bridge, which then cumulate into a big difference. Rather, I am arguing that there is no defensible principled difference between them. My bridge is a chain of inferences, not a slippery slope. (Bridges? Chains? Slopes? I trust the metaphors are stale enough to bear mixing.)

I have now, I believe, said enough to convince you that there is a genuine puzzle here. I will proceed next to suggest what I think is the root cause of the puzzle and then make some suggestions about where a solution is likely to lie. Underlying each view of the Dr. Yu problem is a set of very deeply held, hard-to-shed intuitions—intuitions that, of course, do not square with each other. Let me try to lay bare what they are.

The first set of intuitions consists of our most basic ideas about the attribution of responsibility, the ideas that find their clearest expression in the so-called General Part of the criminal law. These are the rules that tell us when someone is to be blamed for having brought about a certain bad consequence. They require that the wrongdoer have exhibited a suitably

culpable state of mind (intention, knowledge, recklessness, negligence) vis-à-vis the bad consequence, that he have committed an act (as opposed to a mere omission), that the act be proximately (rather than indirectly) connected with the consequence, and that there be no justification or excuse for the act. These, roughly speaking, are the prerequisites for responsibility for bringing about a bad consequence.

We look at Dr. Yu's misconduct and we are inclined to see it as a "consequence" and to see everything that precedes it as the stuff that has to meet the requirements of responsibility before we can blame him. In particular, we are inclined to think that only if the act of swallowing the herb meets the requirements of responsibility are we entitled to blame him for the bad consequence of his misconduct while under the influence of that herb. And, of course, his act of swallowing the herb does not meet those requirements: He has an excellent justifying reason, it seems, for swallowing the herb. This view of the matter, I believe, is what underlies all of the various arguments we are inclined to mount in Dr. Yu's behalf.

There is a second set of intuitions that pushes in precisely the opposite direction. It is an "infinite regress" set of intuitions. It is the thought that once we are willing to bracket someone's misconduct and label it a "consequence," a consequence for which he is to be blamed only if his earlier precipitating conduct meets the usual requirements of responsibility, we can never blame anyone for anything. Whatever someone has done, we will be able to bracket it as a "consequence," focus on earlier conduct, and ask whether it meets the requirements of responsibility. If we do this repeatedly, we are sure to arrive at a moment in time where those requirements are not met. That's the "infinite regress" worry.[11]

The problem is that neither of these intuitions can really be dismissed as entirely fallacious. The first set of intuitions is triggered when we contemplate hypotheticals like those mentioned early on in this section: the soldier who energizes himself by deliberately and momentarily deluding himself into thinking he is shooting at his superior officer, or the surgeon who energizes himself by deluding himself into thinking he is committing an assault. The second set of intuitions is triggered when we contemplate hypotheticals mentioned later in this section: the prosecutor who allows himself to be blackmailed into committing graft; Dr. Yu IV, who does not have the resources to hire bodyguards, and who despite undergoing brain surgery cannot curb his violent impulses enough to prevent himself from committing a murder.

So what is the solution? The solution is that a line will have to be drawn, on one side of which appear the "energized soldier and surgeon" cases, and on the other of which appear the "blackmailed prosecutor" and Yu IV cases.

11. This echoes the choice-character debate. *See* Michael S. Moore, *Choice, Character, and Excuse* and Peter Arenella, *Character, Choice, and Moral Agency, in* CRIME, CULPABILITY, AND REMEDY 29–58 & 59–83 (Ellen Frankel Paul, Fred D. Miller, Jr., & Jeffrey Paul eds., 1990).

Which side of that line the actual Dr. Yu case lies on, we will only be able to determine once that line has been discovered. That's right—the line will have to be discovered, not just made up. I cannot be certain, of course, that our moral intuitions are determinate enough to allow us to discover such a line, but past experience suggests we try, and that there is an excellent chance such a line is just waiting to be discovered. Past experience suggests that lots of moral issues that appear initially indeterminate and insoluble will yield once we do a sufficiently thorough job of canvassing possible "constraining intuitions" in the area.

You might wonder why I am assuming that it will turn out to be a line. Maybe what separates the "soldier" kind of case from the Yu IV kind of case is not a line but a continuum. Maybe the closer a case lies to the "soldier" end of the spectrum the more it should be treated like that case; and the more it lies to the other end the more it should be treated like the Yu IV case. In other words, maybe it is not true that all cases on one side of the line should be treated completely like the soldier case and all cases on the other line should be treated completely like the Yu IV case. That is possible, but unlikely. In ethics, boundaries tend to be lines rather than continua. Even though acts and omissions shade into each other, the treatment of acts and omissions does not shade. We do not say that the closer a piece of conduct is to the omission end of the spectrum, the less it should be treated like an act and the more it should be treated like an omission. Rather, we ask, is it over the line that separates acts from omissions or just shy of it. Why ethical distinctions generally work like that is an interesting mystery in its own right (and one I am in the process of trying to figure out), but it is not peculiarly relevant to the problem at hand.

There is, of course, something very odd and counterintuive about the fact that the soldier and surgeon cases end up calling for a treatment different from the blackmailed-prosecutor and Yu IV cases. That asymmetry is likely to have all kinds of peculiar, interesting, and counterintuitive implications that are still waiting to be discovered and explored.

Meanwhile the original Dr. Yu case stands unsolved. As does that of Arthur McArthur, although it is arguably a somewhat easier case. Arthur McArthur did everything morality required to avoid finding himself in a situation in which he would be too weak to resist temptation, but was—let us suppose—eventually tempted and yielded. That makes him more like Dr. Yu IV than Dr. Yu. And perhaps that means that he in fact should be convicted. But it is hard to be confident of that.

## Punishment Revisited

Let us return to the more traditional kind of preemptive action for which I claimed our exploration of Dr. Yu would have implications. Let us return, that is, to the case of the state that threatens draconian (i.e, highly dis-

proportionate) punishment for what are really "preemptive" offenses. (Offenses like drug dealing or drug use that punish conduct not in and of itself immoral but that has the statistical likelihood of leading to "real" crimes like robbery, arson, and murder.) In a way, a society that threatens such punishment is a bit like Arthur MacArthur. It is arguably doing something perfectly proper—indeed, just exactly what duty calls for—when it threatens wrongdoers with the most draconian punishment. The question arises whether meting out such ostensibly draconian punishment pursuant to a perfectly legitimate threat is thus rendered all right. Some people have argued that it is, some that it is not. Let us examine what they have had to say and then explore the issue further in light of our discussion of Dr. Yu.

Notice that this is different from the question raised about draconian punishment in the previous section. There we asked whether it was moral for a state to inflict some draconian punishment *now* if it thus averted more draconian punishment *later on*. At present, however, we are asking whether it is moral for the state to inflict draconian punishment *now* if it is done pursuant to a legitimate threat *earlier on*. There are basically two positions on this, each represented by one of two seminal papers, the first by Lawrence Alexander, "The Doomsday Machine: Proportionality, Punishment, and Prevention,"[12] the second by Gregory Kavka, "Some Paradoxes of Deterrence,"[13] the former coming out in favor of draconian punishment, the latter against.

Alexander asks the reader to imagine

> a super-sophisticated satellite that can detect all criminal acts and determine the mental state of the actors. . . . If the satellite finds that the actor knew his act was a crime, that he had no recognized excuse or justification for committing it, that he was not acting in the heat of passion or under duress, and that he was not too young, enfeebled, mentally unbalanced, and so forth to be deemed without capacity to commit a crime, the satellite immediately—and without regard to the seriousness of the crime—zaps him with a disintegration ray. Once the satellite detects the crime, it is impossible to prevent punishment of the criminal, no matter how merciful the authorities might feel. The definitions of crimes and the punishments attached thereto can be changed only prospectively. The entire population is informed of the existence of the satellite and what it does.[14]

12. Lawrence Alexander, *The Doomesday Machine: Proportionality, Punishment, and Prevention,* 63 THE MONIST 199–227 (1980).

13. Gregory S. Kavka, *Some Paradoxes of Deterrence, in* Gregory S. Kavka, MORAL PARADOXES OF NUCLEAR DETERRANCE 15–32 (1987). *See also,* Gregory S. Kavka, *A Paradox of Deterrence Revisited, in id.* at 33–56. Particularly important articles in this burgeoning debate are Warren Quinn, *The Right to Threaten and the Right to Punish,* 17 PHIL. & PUB. AFF. 240–47 (1988), and Daniel M. Farrell, *On Threats and Punishments,* 15 SOC. THEORY & PRAC. 125–54 (1989). Other important, relevant entries are cited in Kavka's book.

14. Alexander, *supra* note 12, at 209.

How tolerable would it be to put such a machine in motion? At first glance it appears intolerably cruel. But appearances may be deceiving. Alexander asks us to consider the following:

> Suppose a man receives a phone call from a burglar who says, "I've been spying on you and know you're going out tonight. I plan to burglarize your house in order to steal your valuables. But I want you to know that I have a very bad heart, and if you hide your valuables, I might very well suffer a heart attack by expending a lot of energy and suffering anxiety in looking for them. So please leave them in plain sight; for I am definitely going to enter your house and look for them until I find them or drop dead." The listener hangs up the phone, takes his valuables, hides them on the very top shelf of his closet, and leaves. He returns home and finds the burglar, dead from a heart attack, on the floor. Excessive punishment for a non-violent burglary?[15]

Alexander continues:

> Consider some other examples that I feel are parallel. What if a man keeps a moat to protect his castle (or an electric fence to protect his house), and he receives a letter from someone who says that the first time the castle (house) is deserted he will attempt to enter it; and because he cannot swim (is not shockproof), his death will be on the owner's hands if the moat is not drained (the current not turned off). Is there a duty to drain the moat (shut off the current) in order to avoid excessive punishment? And what if one hides his jewels on top of an unscalable cliff after having been told by the thief that the latter would attempt to climb it if the jewels were placed there.
>
> I might go on in my hypotheticals to drag out vicious dogs, crocodiles and spring guns to protect persons from petty crimes, and pit these devices against petty criminals, whose common denominator is that they all know of the certain consequences of their acts, know that their acts are illegal, are determined to proceed with them anyway, and are acting premeditatedly without any recognized legal excuse, justification, or incapacity.[16]

This is shaping up as a formidable defense of draconian punishment. But so far something critical is missing. Actual draconian punishment would not be carried out by a machine but by human beings. Won't that make a difference? Alexander gives the distinction short shrift. Think of the judge in a regime of harsh punishment, he says, as analogous to a "person returning to his castle to find a forewarned trespasser in the middle of the moat being attacked by crocodiles that the owner can call off by means of a whistle." If having the moat filled with crocodiles is all right, why would it be obligatory to rescue someone caught therein?

But that analogy does not seem totally compelling. Law enforcement officials would seem to be doing substantially more than just letting the

15. *Id.* at 209–210.
16. *Id.* at 210.

wrongdoer die in a deadly moat into which he had fallen by his own actions. To make the case that there is no morally relevant difference between machine-administered punishment and person-administered punishment, a slightly more elaborate argument than Alexander offers seems called for. It would have to go something like this: Alexander might ask us what difference it would make if rather than getting a moat with a crocodile in it, someone had simply gotten himself a ferocious guard dog. Once we say "No difference whatsoever," he might then ask us what difference it would make if instead he had gotten himself a ferocious human guard with instructions to behave like a guard dog (i.e., to kill any intruder). Once we say "No difference whatsoever," Alexander might ask us what difference it would make if instead of always relying on the ferocious guard, we occasionally stepped into his place. After all, if it is all right for us to ask him to kill in our behalf, how can it not be all right for us to do it ourselves? And now we really have arrived at the desired conclusion, a defense of draconian punishment administered by humans, not machines.

The case against Alexander's position is mounted in an article by Gregory Kavka that is ostensibly not about punishment, but about nuclear deterrence. Kavka is concerned with the question whether it is moral to threaten to do something immoral. And he comes to the conclusion that often it is. He thinks of nuclear deterrence as the paradigmatic situation. He pictures a situation in which the United States has been attacked by the Soviet Union and has suffered devastating damage. Any retaliation we now engage in he assumes would be gratuitous cruelty. It would not do anything for us but it would produce the death of a lot of innocent Russian civilians. Yet he also believes that we would be morally safe to threaten to do this very thing which it would be immoral for us to actually carry out. He does not believe, however, that we can infer the morality of retaliating from the morality of threatening to thus retaliate. He draws his chief support from the analogy to rational action. Just as it is possible to rationally render yourself irrational, he says, it is possible to morally render yourself immoral. Nuclear deterrence is such an example.

Presumably Kavka would say the same about punishment. He would argue that threatening to inflict disproportionate punishment might well make good practical and moral sense, but that still does not render the actual implementation of the threat moral. What about the argument that juxtaposes the doomsday machine with a guard and the guard with the person being guarded? Kavka would almost certainly approve of the creation of the doomsday machine, for the same reason that he would have no problem creating a doomsday machine that was certain to retaliate against attacking Russians. He would even approve of the hiring of a guard dog or even the replacement of the guard dog with human beings. Finally, Kavka would even be willing to say that the person who makes arrangements to play the part of the guard, should the need arise, is acting morally. It is just

when the actual need arises that he is no longer acting morally, and neither is his guard.

So who is right, Alexander or Kavka? I wish I knew. The case turns out to be very similar to that of Dr. Yu, and therefore many of the arguments—nearly all of them, in fact—would apply to this debate. They help to make it even more inconclusive than it already is, but they also, I think, illuminate it.

Let us take a closer look at those arguments and see how they play out.

*First.* I argued in Dr. Yu's behalf that when he tried to kill that child, the context should be permitted to wash away his guilt in the same way it is permitted to do in a case of self-defense. The context here is one in which all of society adopted a certain draconian punishment regime. Law enforcement officials are simply acting on society's orders. It seems natural to say that the morality of their actions should depend on the morality of the orders they received. Because those orders were moral when they were issued, namely before the crime was committed, then the carrying out of those orders is moral.

One might try to make the analogy to self-defense even tighter by thinking of someone acting in self-defense as giving himself a mental order at time 1, when he notices he is being attacked, to kill his attacker at time 2. We can then think of him at time 2 as simply following the orders he gave himself at time 1. Now it would seem as though his actions at time 2 are moral precisely if the order he gave himself at time 1 was moral.

So much for the argument *for* draconian punishment. If we turn to the argument I made in reply when discussing Dr. Yu, a strong objection to draconian punishment emerges: Maybe we should not think of draconian punishment as analogous to self-defense but rather as analogous to the case of the blackmailed prosecutor—the prosecutor who allows an ex-convict to make an embarrassing movie about himself in return for obtaining the release of the prosecutor's children, and who is later blackmailed into committing various property crimes on pains of having the movie released. This is a case in which we are tempted to not pay much attention to the perfectly honorable way in which the wrongdoer, the prosecutor, caused himself to do dishonorable things later on. Perhaps draconian punishment is like that: Passing draconian legislation is a perfectly honorable way in which we cause ourselves to do dishonorable things later on.

*Second.* The abandonment argument I made in Dr. Yu's behalf would not seem to carry over very easily to the problem of draconian punishment. The abandonment defense is generally only available to defendants who never actually bring about any harm. That was true of Dr. Yu. But it is not true of law enforcement officials in a draconian regime: They are inflicting harm when they punish people. It is not crystal clear, however, that the abandonment defense should only be available to defendants who fail to bring about harm. Let us imagine that in Dr. Yu's case someone had secretly bribed Yu's

bodyguards to not intervene when he grows violent, and that as a result he had killed that child. Would that alone really deprive him of his abandonment argument? Arguably it should not. As long as Yu tried very hard to make sure that his actions while under the influence of the herb harmed no one, many would say he is the moral equivalent of someone who attempts a crime, abandons it—but whose victim nevertheless dies because an unrelated assassin suddenly appears on the scene and kills him. We might look at law enforcement officials in a draconian regime in the same way. They tried very hard to avert any harm coming from the adoption of draconian laws: They advertised to everyone around just how high a price wrongdoers would have to pay for violating the law. If, nonetheless, someone comes to harm, that is because the offenders essentially behaved like those delinquent bodyguards and caused the best-laid plans to come to naught.

That said, there is no doubt that the abandonment argument works much better in behalf of Dr. Yu than in behalf of draconian punishment.

*Third.* My epileptic seizure analogy also works a bit better for Dr. Yu than for draconian punishment, but it is not a hopeless line of argument either. Given that the police, the judges, the prison wardens are all just playing out a small part in a gigantic script, they seem in some ways as nonresponsible for what they are doing as is an epileptic. It would seem that the wrongness of what they are doing should thus hinge, as in an epileptic's case, on what caused them to act as they did. Because we know that the adoption of such laws, the causing of such behavior, was all right, perhaps they too can escape reproach?

*Fourth.* We come to what I consider the strongest argument both in behalf of Dr. Yu and in behalf of draconian punishment: the example of the soldier who likes to energize himself by imagining the enemy he is firing at to be a superior officer he particularly despises. Although this example strongly parallels Dr. Yu's case, it seems at first very different from the draconian punishment scenario. The key difference would seem to be that the soldier does not actually inflict any harm. He kills someone, to be sure, but someone he is supposed to kill. Draconian punishment involves the harming of people who do not deserve such harsh treatment. But we can change the soldier example slightly to eliminate that difference. We can imagine that in the brief moments while the soldier diverts his gaze from his target, so as to better immerse himself in his self-delusion about who he is shooting at, the enemy soldier manages to run away, and by a near-miraculous accident the soldier's hated superior happens to step into the very spot the enemy soldier was previously occupying—and gets shot. Intuitively, it would seem the soldier who shoots him would still deserve to be acquitted. Here now we have a case in which someone morally causes himself to engage in conduct that is both highly immoral and harmful but who nevertheless seems to deserve an acquittal. Which is exactly the analytical structure of the draconian punishment problem.

Of course, there are competing analogies on the other side, like the case of Dr. Yu IV, who is born with an angry temperament, which he unsuccessfully tries to cure with psychotherapy, electroshocks, and brain surgery. When he eventually kills someone, we will not be inclined to acquit him by reason of his diligent efforts to keep his demons in check. Arguably, this is how we should think of draconian punishment: We could think of the adoption and aggressive advertisement of draconian penalties as paralleling Yu IV's efforts to avert his own wrongdoing by psychotherapy, electroshocks, and brain surgery. In which case draconian punishment does not escape the charge of immorality.

My aim in revisiting the four sets of arguments pro and contra Dr. Yu has been to show that draconian punishment, like the case of Dr. Yu, is shrouded in moral uncertainty. Like the case of Dr. Yu, it belongs to that large family of cases in which we cannot decide whether to "bracket" someone's misconduct as a consequence and then investigate its antecedents or whether we should forget about its antecedents and focus on the misconduct itself. Some of those cases, I suggested, should be treated one way, and some the other. What the appropriate way is for thinking about draconian punishment is as yet undetermined.

## IV. CONCLUSION

Sometimes it is possible to prevent oneself from committing a serious wrong in the future by committing a less serious wrong right now. The central questions taken up in this essay concern whether one is obligated to do this; whether one is entitled to do this; and what implications the answers to these questions hold for the more familiar issue of whether one may forestall ANOTHER person's wrongdoing with wrongdoing of one's own (by engaging, for instance, in the preventive detention of innocents who have a mere statistical likelihood of becoming criminals).

The answers to all of these questions are fraught with paradox. It turns out, for instance, that the person who forestalls his own future wrongdoing with less serious present wrongdoing will often manage both to improve his moral position and to merit condemnation at the same time. (This is what I call the "gateway sin" paradox.) It also turns out that the person who refuses to prevent his own future wrongdoing with present wrongdoing thereby often manages to cleanse his subsequent wrongdoing of all moral taint. (That is the problem of Dr. Yu.) Finally, all of this turns out to imply that controversial practices like preventive detention, preemptive self-defense, or draconian penalties are morally much more defensible than usually thought.