Contrived Defenses and Deterrent Threats: Two Facets of One Problem

Claire Finkelstein* & Leo Katz**

What relation do the various parts of a plan bear to the overall aim of the plan? In this essay we consider this question in the context of two very different problems in the criminal law. The first, known in the German criminal law literature as the Actio Libera in Causa*, involves defendants who contrive to commit crimes under conditions that would normally afford them a justification or excuse. The question is whether such defendants should be allowed to claim the defense when the defense is itself either contrived or anticipated in advance. The second is what we call the question of deterrent threats: is it permissible to threaten to do more than it is actually permissible to do, in order to deter the wrongdoing of another? Furthermore, if it is permissible to issue such a threat, does it then become permissible to follow through on it when the threat fails to deter, thus rendering permissible an action that would otherwise be impermissible? These two problems—the problem of contrived defenses and the problem of deterrent threats—appear to be mirror images of one other. The first involves a morally permissible act embedded in an immoral course of conduct, while the second involves a morally impermissible act embedded in a moral course of conduct. The question we raise is whether the larger plan or course of conduct should help to determine the character of the individual acts that constitute that plan, and whether there is a consistent approach to plans and their component parts that provides plausible answers to legal questions of both sorts.*

I. INTRODUCTION

What is the relationship between a complex whole and the individual parts that constitute it? In particular, when do the characteristics of the larger entity carry over to the characteristics of its components? In the physical world, the answer is—rarely.

* Professor of Law and Philosophy, University of Pennsylvania.

** Frank Carano Professor of Law, University of Pennsylvania. We are grateful to Larry Alexander, Pete Alces, Sharon Byrd, William Ewald, Jim Gordley, Michael Moore, Stephen Morse, and Paul Robinson, as well as to our colleagues at the 2005 University of Pennsylvania Law School faculty retreat and to members of the audience at the William and Mary Conference on Law and Morality where an earlier version of this paper was presented. Each of us is indebted to Sandy Kadish, for his friendship, his mentorship, and for teaching us, in different ways, the profound interest that comes from thinking about ordinary problems in criminal law through the lens of philosophical reflection.

The molecules of a blue object are not themselves blue; the molecules of liquids are not themselves wet; the sides of triangles are not themselves triangular, and so on. But is it the same with complex normative properties?[1]

We will focus on situations in which someone makes a plan involving several different steps. On the one hand, we could ask whether the plan, taken as a whole, is morally or legally permissible. On the other hand, we could ask about each action required by the plan, whether *it* is morally or legally permissible. The problem that concerns us is whether the answer we give to the first question is necessarily the same as the answer we give to the second question: If the plan as a whole is morally (or legally) permissible, should we expect the same to be true of all its subparts? Conversely, if all the subparts of a plan are morally (or legally) acceptable, should we expect the same to hold of the overall plan?

We will address this topic in two different contexts. The first is the problem of defendants who manufacture the conditions of their own defense, which German criminal law scholars have dubbed the problem of the *Actio Libera in Causa*.[2] The standard example is the case of the defendant who deliberately drinks himself into a state of irresponsibility in order to commit a crime while in that state. The question raised by such cases is whether the fact that the defendant's action is part of an overall plan should lead one to evaluate the action differently from the way one would if one were to consider it in isolation. In particular, should we deny the defendant a defense we would otherwise grant because the defense is claimed in the context of a plan to produce it?[3]

---

[1] G.E. Moore famously argued that complex normative entities often form what he called "organic unities," in which the quality attributable to the entity as a whole could not in turn be attributed to its component subparts. G. E. MOORE, PRINCIPIA ETHICA ch. 1, § 18 (Cambridge Univ. Press 1903). In effect, the question we are raising is to what extent a plan forms an organic unity with respect to characteristics like permissibility or rationality relative to its constituent elements.

[2] *See* MICHAEL HETTINGER, DIE „ACTIO LIBERA IN CAUSA": STRAFBARKEIT WEGEN BEGEHUNGSTAT TROTZ SCHULDUNFAEHIGKEIT? EINE HISTORISCH-DOGMATISCHE UNTERSUCHING (1988); JOACHIM HRUSCHKA, STRAFRECHT NACH LOGISCH-ANALYTISCHER METHODE (1983); U. NEUMANN, ZURECHNUNG UND VORVERSCHULDEN (1985); DOROTHEE SYDOW, DIE ACTIO LIBERA IN CAUSA NACH DEM RECHTSPRECHUNGSWANDEL DES BUNDESGERICHTSHOFS (2002); HUBERT STÜHLER, DIE ACTIO LIBERA IN CAUSA DE LEGE LATA UND DE LEGE FERENDA: EINE ANALYSE VON RECHTSPRECHUNG UND LITERATUR VERBUNDEN MIT EINEM GESETZGEBUNGSVORSCHLAG (1999); RENÉ ZENKER, ACTIO LIBERA IN CAUSA: EIN PARADOXON ALS ÖFFENTLICHER STRAFANSPRUCH IN EINEM VOM SCHULPRINZIP GEPRÄGTEN RECHTSSTAAT (2003). The problem has received little attention in Anglo-American criminal law theory, with the notable exception of MIRIAM GUR-ARYE, THE ACTIO LIBERA IN CAUSA (1984); MICHAEL S. MOORE, ACT AND CRIME: THE PHILOSOPHY OF ACTION AND ITS IMPLICATIONS FOR CRIMINAL LAW 35–36 (1993); Paul H. Robinson, *Causing the Conditions of One's Own Defense: A Study in the Limits of Theory in Criminal Law Doctrine*, 71 VA. L. REV. 1, 3 n.6 (1985); *see also* Claire Finkelstein, *Involuntary Crimes, Voluntarily Committed, in* CRIMINAL LAW THEORY: DOCTRINES OF THE GENERAL PART 147 (Stephen Shute & A.P. Simister eds., 2002); LEO KATZ, ILL-GOTTEN GAINS, at pt.I (1996).

[3] A related problem is raised by defendants who are aware of, but do not contrive, the conditions of their own defense, which the Germans have dubbed the *Actio Illicita in Causa*. The difference between the *Actio Libera in Causa* and the *Actio Illicita in Causa* will not be important for our purposes, so we will treat such cases together.

The mirror image of the *Actio Libera in Causa* is raised by what we shall call the problem of deterrent threats. Is it permissible for a person to threaten to use more force than it is in fact permissible for him to use? And if so, does the issuance of a sincere deterrent threat which *fails* to deter make it in turn permissible to execute the threat so issued? In the criminal law, the question can be raised in the context of the use of deadly force to protect property. While it is normally not permissible to use deadly force merely to protect property, it may be permissible to threaten to use such force as a way of deterring potential wrongdoers from theft or other property crimes. The issue has also been raised in other contexts, notably in the literature discussing the morality (and rationality) of the United States policy of nuclear deterrence during the Cold War.[4] Against the policy of Mutually Assured Destruction, the argument was often made that it could not be moral to issue a threat on which it would not be moral to act. And since surely it would not be moral to launch a weapon of mass destruction, it could not be moral to threaten to do so either. If, however, it makes sense to evaluate actions that occur in the context of plans differently from those same actions when they are self-standing, then issuing and eventually following through on deterrent threats of the above sort may look more acceptable than it otherwise would.

While our two contexts are quite different, they display a surprising symmetry. The *Actio Libera* involves seemingly innocent, or at least excusable, actions embedded in an overall wicked plan. And the question we must ask is whether the plan's wickedness should be imputed to the otherwise innocent conduct to *inculpate* the defendant, who otherwise would have had an excuse or justification. The logic of deterrent threats, by contrast, involves seemingly immoral conduct embedded in an overall *moral*, or at least permissible, plan. The question in this case is whether the plan's justifiable character should be imputed to the otherwise immoral conduct to *exculpate* the defendant, who otherwise would have been guilty of a crime.[5] Both problems raise the question of whether the larger plan should play a role in assessing the individual actions that help to constitute it. Exploring these two types of plans—contrived defenses and deterrent threats—may not provide a definitive answer to the problem of the relation between plans and their component parts, but it should at least allow us to trace a set of consistent responses to cases of this sort.

In Part II of this Essay, we will sketch the *Actio Libera* problem. While this problem has been extensively written about elsewhere, our aim here is merely to summarize and clarify the existing debate and to indicate reasons for favoring one approach over others. In Part III, we will turn to the problem of deterrent threats. As in the case of the *Actio Libera in Causa*, we believe there is reason to favor one approach to the morality of deterrent threats over others. In Part IV, we will attempt to articulate the parallels between the *Actio Libera in Causa* and the deterrent threat situations in greater detail. In so doing, we hope to show that the favored solution to

---

[4] *See, e.g.*, GREGORY S. KAVKA, MORAL PARADOXES OF NUCLEAR DETERRENCE 16–21 (1987).

[5] Some aspects of the relationship between these kinds of cases were discussed in Leo Katz, *Pre-empting Oneself: The Right and the Duty to Forestall One's Own Wrongdoing*, 5 LEG. THEORY 339 (1999).

the *Actio Libera in Causa* problem receives support from the plan-based approach we sketch in the context of deterrent threats, and that the reverse is also true. We conclude in Part V.

II. THE *ACTIO LIBERA IN CAUSA*

The *Actio Libera in Causa* is the name German criminal law scholars have given to situations in which a defendant arranges to commit a crime by rendering himself mentally irresponsible.[6] In addition to the defendant who drinks in order to carry out a crime while drunk, examples might include: the person who has himself hypnotized so that he will murder his wife while in a trance; the defendant who intentionally places himself in a situation in which he knows he will be coerced into committing a crime he otherwise wants to commit; or, more colorfully, the defendant who has himself shot out of the mouth of a cannon and into the window of a jewelry store he intends to burgle. The doctrine has also been generalized to justification defenses, such as the person who provokes another to attack him in order to be able to kill the latter in self-defense, or the person who sets a forest fire so that he can later claim a justification for burning someone else's property to create a firebreak.[7]

A related doctrine, known as the *Actio Illicita in Causa*, encompasses defendants who perform actions knowing, rather than intending, that they are likely to lapse into an irresponsible state.[8] Examples include the defendant who drives knowing he is subject to epileptic seizures, or the person who exposes himself to his enemy knowing he is likely to lose control and attack him. These defendants create, as Paul Robinson aptly puts it, "the conditions of [their] own defense."[9] The question is whether defendants in this position should be permitted to avail themselves of the defense they have thus anticipated. In what follows, we will consider three approaches that have been offered to this problem. The first of these solutions has thus far dominated the literature, and thus we consider it most extensively.

A. *The Traditional Solution*

The traditional solution to the *Actio Libera* (or *Actio Illicita*) *in Causa* problem would deny the strategically intoxicated or strategically provoking defendant his defense.[10] The argument for denying the defense is most powerful where result crimes are concerned. If we are dealing with a homicide, for example, the defendant's action of getting drunk inaugurates a chain of events culminating in the victim's death. The original action of getting drunk takes the place of pulling the trigger of a gun pointed at the victim, and because at that very moment the defendant is still sober

---

6 Finkelstein, *supra* note 2, at 147.

7 *See* Robinson, *supra* note 2, at 3–4.

8 Finkelstein, *supra* note 2, at 147 n.10; Robinson, *supra* note 2, at 37–38.

9 Robinson, *supra* note 2, at 2.

10 *Id.* at 4–8.

and fully responsible, we have no problem holding him liable for murder.[11] The defendant himself is like a "bullet in flight," and though he must still engage in a series of involuntary bodily movements in order to carry out the crime, he is fully responsible for those movements, given that they were caused by his own earlier act of getting drunk.

Michael Moore explains the traditional approach thus:

> [I]f, from the big bang that apparently began this show to the heat death of the universe that will end it, the court can find a voluntary act by the defendant, accompanied *at that time* by whatever culpable mens rea that is required, which act in fact and proximately causes some legally prohibited state of affairs, then the defendant is prima facie liable for that legal harm . . . . If there is *any* point in time where the act and *mens rea* requirements are simultaneously satisfied, and from which the requisite causal relations exist to some legally prohibited state of affairs, then the defendant is prima facie liable.[12]

The same point can be extended to cases of contrived justification defenses. If the defendant intentionally provoked his attacker in order to be able to attack him in supposed self-defense, the defendant caused his own need for self-defense by that earlier act. Because the defendant possessed the requisite mens rea at the earlier moment at which he voluntarily and intentionally caused himself to attack, he is liable for the attack, and the claim of self-defense should be unavailable to him. In the German criminal law literature, this is known as the *Vorverlegungstheorie* (the "prepositioning theory") or the *Tatbestandsmodell* (the "elements of the crime model").[13]

For the most part, American criminal codes display an unsystematic version of the traditional approach. In the Model Penal Code, for example, the problem is not conceived as a whole. Rather, there is periodic mention of a contriving or aware defendant who creates the conditions of his own defense. The Model Penal Code's self-defense provision denies the actor the use of lethal force if he provoked the use of force against himself in the same encounter.[14] The duress provision denies the defendant the defense if he placed himself in a position in which he knew he might be subjected to duress.[15] The Code's account of necessity is similar and provides that the defendant foregoes the defense of necessity if he was negligent or reckless in creating the situation that required him to violate the law.[16] The defendant can then be

---

[11] *Id.* at 7.

[12] MOORE, *supra* note 2, at 35–36.

[13] *See* sources cited *supra* note 2.

[14] MODEL PENAL CODE § 3.04(2)(b)(i) (1985).

[15] *Id.* § 2.09(2).

[16] *Id.* § 3.02(2).

convicted of a related crime for which the mens rea of negligence or recklessness suffices.

While the traditional approach seems a sensible one, over time many difficulties with it have emerged.[17] Although none of the objections deals the approach a mortal blow, together they significantly diminish its attractiveness. We will summarize these difficulties below.[18]

The first problem with the traditional approach appears when one attempts to apply that solution to conduct crimes. Consider a crime like burglary, which requires the defendant to have "enter[ed] a building . . . with purpose to commit a crime therein."[19] Can the defendant be said to satisfy this act definition in the case in which he had himself shot from the mouth of a cannon? Although he did enter a building with intent to commit a crime, he did not do so *voluntarily*, since he was in an irresponsible state at the time. What about the fact that he was in a fully responsible state when he arranged to have himself hypnotized in order to commit a burglary, as the traditional approach would have it? The problem with this earlier moment is that he did not enter a building at that time, since getting oneself hypnotized is not itself entering a building. So the trouble is that at the earlier moment in time, T_1, the defendant was responsible but he was not entering a building, and at the later moment, T_2, when the defendant *was* entering a building, he was not in a responsible state. To put the problem more technically, it looks as though there is no concurrency of act and mental state, which is required for the defendant to be liable for a crime.

The problem, then, with conduct crimes is that they require a particular action by the defendant, such as driving,[20] having intercourse,[21] or breaking into a dwelling.[22] The importance of the defendant's satisfying a specific conduct element for the latter group of crimes means that a defendant who manufactures the conditions of his own defense cannot be deemed to satisfy that element by virtue of the earlier act by which he caused himself to perform the later illegal act. At the earlier moment in time, the defendant is at most *causing* himself to do those things, which would be helpful for liability if the crime were a result crime. But for a specific conduct offense, the defendant is unlikely to meet the actus reus requirement on the basis of the earlier act.

Traditionalists tend to gloss over the definitional problem, treating it as a mere drafting matter. Michael Moore, for example, suggests that there is no reason to cleave to "stereotypes" about how criminal acts are performed: "Just as one kills by causing death, so one rapes by causing penetration, one hits by causing contact, one maims by causing disfigurement, and one takes by causing movement of the object

---

[17] *See* Paul H. Robinson et al., *Making Criminal Codes Functional: A Code of Conduct and a Code of Adjudication*, 86 J. CRIM. L. & CRIMINOLOGY 304, 324–27 (1996).

[18] These difficulties with the traditional approach have been most clearly articulated in the German criminal law commentary. *See* sources cited *supra* note 2.

[19] MODEL PENAL CODE § 221.1(1) (1985).

[20] *Id.* § 223.9.

[21] *Id.* § 213.0.

[22] *Id.* § 221.1.

taken."[23] As a point about ordinary language, Moore is certainly right when it comes to verbs like "killing."[24] For the fact that a person does something at T_1 that later causes someone's death at a wholly different time, T_2, does not in any way interfere with our ability to say that what is done at T_1 is an instance of killing. If the defendant places poison in his wife's tea at T_1, and his wife later drinks the tea and at T_2 dies, the defendant's action at T_1 is an instance of killing, even though his wife did not die until T_2.[25] This suggests that any number of actions performed at T_1 can be "killings." There is no stereotypical killing, and the same can be said for "causing death."

But the same point simply does not hold for conduct crimes. A defendant does not have an infinite number of ways to "unlawfully enter or remain in a building," or "unlawfully remove property," or "alter a writing with intent to deceive," or "have intercourse with a woman without her consent," and so on. The question is how much flexibility there is in the conduct requirement in each case. Must a person physically place his body inside the building to enter it? Or can he be constructively deemed to have entered it by seeing a hypnotist, drinking a potion, or getting himself shot from a cannon? To us it seems to strain ordinary meanings to say that a person is entering a building just because he is doing something that will later *cause* himself to enter a building, especially when the thing he is doing is an action as different from entering a building as visiting a hypnotist. Sandy Kadish has suggested that conduct crimes have the special feature that they are "nonproxyable," meaning that they cannot be accomplished through the intervention of another. This is another expression of the fact that to cause a rape is not to rape, to cause a burglary is not to burglarize, and to cause an assault is not oneself to assault. This is a matter, not only of ordinary language, but of morality as well.[26]

. . .

---

[23] Michael Moore, *Causation and Responsibility*, 16 SOC. PHIL. & POL'Y 1, 39 (1999).

[24] *Id.*

[25] For a detailed discussion of this problem, see Judith Jarvis Thomson, *The Time of a Killing*, 68 J. PHIL. 115 (1971).

[26] Kadish uses this feature of conduct crimes to explain the need for a law of complicity. *See* Sanford Kadish, *Complicity, Cause and Blame: A Study in the Interpretation of Doctrine*, 73 CAL. L. REV. 323, 373 (1985).

III. THE PROBLEM OF DETERRENT THREATS

A problem we will somewhat imprecisely call the problem of deterrent threats has repeatedly engaged the attention of moral philosophers and criminal law scholars. The central question it addresses is whether it is permissible to threaten to use more force to deter another's wrongdoing than it is permissible actually to use. Let us consider a common situation in which this problem might arise.

Suppose a man is defending his property against a thief. What sort of force is the man permitted to use? The answer is that he may use whatever force is necessary to prevent the wrong, but within sharply defined limits. He may not generally use *deadly* force merely to protect property, unless his person, or his domicile are also under attack.[49] While this limitation on the use of deadly force might be debated, we will take the restriction for granted, at least in the first instance.

There are three different questions that arise in this context, all of which we will consider as part and parcel of the problem of deterrent threats. The first is as follows: even though it is not permissible for the property owner to use deadly force in defense of his property, would it nevertheless be permissible for him to *threaten* to use deadly

---

[49] *See* MODEL PENAL CODE § 3.06(3)(d) (1985).

force? May he, for example, brandish a weapon and say: "If you don't drop that television set, I will shoot"? The natural, and most widely maintained view, is that he is allowed to threaten to use more force than he is allowed to use in this context. Threatening to use deadly force is not equivalent to using it.[50] This position is clearly defensible in the case of a bluffing threat. If, for example, the property owner has no intention of actually *using* such force, then surely he is allowed to bluff in order to deter a potential malefactor from committing his misdeed.

But suppose he is *not* bluffing, and that his threat to use deadly force should his threat fail to deter is sincere. Is he permitted sincerely to threaten to do what he is not permitted actually to do? One is tempted to extend the answer from the insincere threat: A threat is not the same as the use of actual force, and while lethal force to protect property is surely disproportionate, a deterrent threat to use such force arguably is not. The prohibition in this case is not on the *intention* the property owner forms to use lethal force, but rather on the actual use of such force. While there is admittedly a problem with avoiding the use of force once a sincere threat to use it has been issued, that is arguably more of a causal problem than a moral one. If issuing a sincere threat to use deadly force is *possible*, in the face of the knowledge that actually using such force is not permissible, then issuing such a threat is arguably permissible.

Moreover, notice that issuing a threat is particularly justifiable if the threat is likely to be effective, since that will reduce the probability of having to follow through on the threat, given that the property owner thereby minimizes the probability of wrongdoing on the part of others. Thus threatening as a form of prevention, even threatening to do what it is not permissible to do, is arguably justifiable as a minimally invasive way of defending against rights violations.

On the other hand, many philosophers have taken the position that it is never permissible sincerely to threaten to do that which it is not permissible actually to do.[51] The moral quality of the threat, they think, depends on the act actually threatened. This position may receive support from the law of attempt, for arguably the property owner could be found guilty of an attempt on the basis of his threat alone. If *conditionally* intending to use deadly force is the equivalent of intending to use it,[52] and if brandishing a weapon counts as a substantial step in the direction of using such force (as it arguably does), then the person threatening to use deadly force would be guilty of attempted murder. Let us call the position that the permissibility of the threat depends on the permissibility of the act threatened the Backward Induction View.

Now there are reasonable objections to the Backward Induction View. First, there is much support for the view that conditionally intending to do something is not

---

[50] The Model Penal Code specifically excludes brandishing a weapon from the definition of deadly force. *See* MODEL PENAL CODE § 3.11(2) (1985) ("A threat to cause death or serious bodily injury, by the production of a weapon or otherwise, so long as the actor's purpose is limited to creating an apprehension that he will use deadly force if necessary, does not constitute deadly force.").

[51] *See, e.g.*, ANTHONY KENNY, THE LOGIC OF DETERRENCE (1985).

[52] *See* MODEL PENAL CODE § 2.02(6) (1985) ("When a particular purpose is an element of an offense, the element is established although such purpose is conditional, unless the condition negatives the harm or evil sought to be prevented by the law defining the offense.").

the same as intending to do it. Second, on the attempt argument, there is arguably a problem in the low probability cases with substantial step liability: If brandishing a weapon were so inherently unlikely to result in using that weapon, the threat perhaps should not count as a substantial step in the direction of an actual killing. Arguably, to use the Model Penal Code's language, such conduct would not then be "strongly corroborative of the actor's criminal purpose."[53] From a broader moral point of view, one can say that the defendant is engaging in an act that has an ostensibly high probability of a favorable outcome—deterring a theft—and a very small risk of a deadly outcome. This makes it not too different from other risky forms of defensive measures deemed acceptable, such as allowing police officers to bear arms. Once again, the more effective the threat, the more likely it is to deter another's wrongdoing, and thus the further it will be from the actual use of deadly force, and the closer to a non-aggressive means of protecting one's property.

Now let us assume that it is in fact permissible to threaten more force than it is permissible to use. This makes it possible to ask a further question: Would it be permissible to install a mechanical device, such as a spring gun, as a way of making such a threat credible? On the one hand, it seems difficult to see how such a device could be legitimate, if the action of the spring gun would not be permissible were it performed by a human agent. Presumably, if the use of deadly force to protect property is impermissible, then using a spring gun that would automatically inflict the same force would not be permissible either. There are, however, several interesting arguments in favor of the permissibility of spring guns or comparable mechanical devices in this context.

First, it is still possible to argue that like the threat, the use of the device does not itself constitute deadly force. Deadly force is so-named not merely because there is *some* possibility of death, but because there is a *significant* possibility that someone might die, combined with an intent on the part of the agent to inflict lethal harm or at least serious bodily injury. If setting up a spring gun has deterrent benefit, it may once again increase the effectiveness of the threat, which will *minimize* the likelihood of actually having to use deadly force. A threat to use deadly force, backed up by a mechanical device, may also be seen as an attempt to avoid the use of actual deadly force. And if the issuance of a sincere threat to use deadly force is permissible, then the issuance of such a threat, aided by an automatic retaliation device, is also permissible.

An even stronger argument in favor of the permissibility of spring guns and their ilk is suggested by Larry Alexander in his celebrated article on doomsday machines.[54] Imagine a property owner who decides to hide his property in some dangerous spot, a mountain top, a cave with wild animals in it, a hole full of snakes, the top of a hard to reach armoire, etc. The thief calls up the owner and tells him that he is determined to get to the property or die trying. Does the owner have a duty to move the property to

---

[53] MODEL PENAL CODE § 5.01(2) (1985).

[54] *See* Lawrence Alexander, *The Doomsday Machine: Proportionality, Punishment and Prevention*, 63 MONIST 199 (1980).

a more accessible place? And if not, as he surely does not, how is moving the property to a dangerous location morally different from using a spring gun to create a similarly dangerous situation? If the property owner knows that the thief will keep trying to steal his property to the bitter end, does that not make the property owner as responsible for the thief's death as he would be in the case of a spring gun? Put another way, if it is not impermissible to protect one's property by placing it on a high mountain top where the air is thin, or to place it where it is protected by dangerous animals, then surely it is not impermissible to protect it with a spring gun, provided, once again, that potential thieves are clearly informed of the presence of the gun. Alexander argues accordingly that the use of a mechanical device to protect one's property is permissible, as long as the potential thief is on notice of the presence of the spring gun.[55]

Let us, then, provisionally accept the conclusion that it is permissible to protect one's property by use of a spring gun or comparable dangerous device. We come then to a third issue. Suppose that the homeowner were to arm himself with a "spring" gun that required manual activation for purpose of protecting his property with deadly force. This is, of course, precisely the move that courts and criminal codes regard as impermissible. But notice: If it is permissible to threaten to use deadly force to protect property, and permissible to set up an automatic execution device to make such a threat credible, how could it be impermissible for a homeowner to use *manual* force to do the same? In other words, if the threat to use force is permissible, clearly advertised, and sincerely meant, then an actual shooting ought to be permissible in case someone actually attempts to steal the television, whether that shooting is a mechanized response to a wrongful act or a manual one. But if this is correct, we have bootstrapped ourselves into the conclusion that using deadly force to protect property is permissible, despite our firm assumption at the outset that it is not.[56] Has something gone wrong?[57]

One rather suspects something has, yet it is difficult to identify the precise flaw in the argument. To recap, the first crucial step was to say that it may be permissible to defend property with a *threat* to use deadly force, under circumstances in which it would not be permissible actually to use it. The second was to say that if the threat was sincerely issued, and was permissible *qua threat*, then it should be permissible to set up a device, such as a spring gun, that would automatically carry out the threat. And the third was to say that if it is permissible to use an automatic retaliation device to make good on a sincerely issued, legitimate, threat to use force, then it should be permissible to do the same without the automatic execution device but manually

---

[55] *Id.* at 216. Strictly speaking, it is not clear why Alexander thinks the point depends on notice, since surely one is not obligated to put the thief on notice before stashing one's jewels on a remote mountaintop. But admittedly the implications of eliminating notice would be particularly far-reaching, since it would suggest that the state may punish preventive action to forestall another's wrongdoing without even warning potential malefactors of the possible consequences of their behavior.

[56] *Id.*

[57] For further discussion of the nature of deterrent threats in the law-enforcement context, see Claire Finkelstein, *Threats and Preemptive Practices*, 5 LEGAL THEORY 311 (1999).

instead. It was in this way that we moved from the permissibility of threatening more than it is permissible to *do*, to the permissibility of doing the very thing assumed to be impermissible. Which step might we best reject?

First, we could reject the suggestion that it is permissible to issue a sincere threat to use more force than it is, prima facie, permissible to use. Against this position, however, lies the compelling intuition that *threatening* to use force is simply not the same as using force. And further, as we have seen, *threatening* to kill may be the least invasive way to forestall another's wrongdoing, and thus arguably it is to be preferred as compared with more severe invasions of personal freedom and bodily integrity, such as attempting to deter future wrongdoers by punishing strenuously after the fact. It seems difficult to reject this step.

Second, we could accept the suggestion that it is permissible to threaten to do more than it is permissible actually to do, but reject the permissibility of installing an automatic device as a way of enforcing the threat. That would require us, among other things, to distinguish the installation of such devices from merely hiding one's property in a dangerous place and keeping it there even though the thief has informed us that he will keep searching until he finds it or is killed in the process. As we have seen, this is not an easy distinction to defend.

Third, we could reject the move from automatic threat execution to manual threat execution. That is, someone might say it is permissible to set up a spring gun or other automatic device to make good on a deterrent threat, and permissible therefore for that device to fire should the threat fail to deter, but that this is entirely different from choosing to follow up on a deterrent threat with a separate, intentional act. The legitimacy of the threat—justified as it is by the fact that it stands a high likelihood of deterring the wrongful act with a low probability of having to inflict actual harm—serves to justify only *one* act, and in many cases that single, emphatic act of threatening is worth the costs. In other words, the cost-benefit analysis that allows us to justify threatening to use deadly force also justifies the entire set-up (threat and automatic execution device), when the two can both be justified in a single, up-front act. In the case of manual threat execution, by contrast, the second act—pulling the trigger of the gun when the threat fails to deter—must be justified only *after* the justification for issuing the threat has passed. We know at this second stage that the threat has failed to deter, and there is nothing further to be gained from actually making good on the threat. So arguably the case is entirely different when the threat must be executed with human deliberation than when it can be executed automatically. *Now* if the threat fails to deter, the agent must reflect on whether to pull the trigger, and there seems to be nothing one can say to justify pulling the trigger at this point. Since no deterrent or preventive rationale can apply, the only justification would be retributive, and it is clearly not retributively permissible to inflict the death penalty for theft.

There are, however, serious objections that can be leveled at this third attempt to avoid the bootstrapping argument. Most compelling, perhaps, is the fact that the line between spring guns and manual use of lethal force seems a highly artificial one. While the one act/two act distinction carries some plausibility, it is not hard to

imagine that an agent can make his own actions nearly as automatic as the actions of a spring gun, simply by refusing to reconsider his own plan. Alternatively put, the "plan" of issuing the deterrent threat and then making good on it should the threat fail to deter provides the same glue between the earlier act of threatening and the later act of shooting as the automatic mechanism of the spring gun. There is a rich literature in philosophical rational choice theory on the topic of so-called "resolute choice." One of us has defended a version of the resolute view in the context of debates about the nature of rationality and planning agency.[58] While this is not the place to recapitulate all the arguments in favor of resolute choice in the rationality literature, it should suffice to notice the *moral* implausibility of distinguishing pre-commitment from resolution in the context of threat execution. If a human being can make himself into a kind of automatically firing spring gun by steeling himself for the task and refusing to reconsider should his threat fail to deter, then his shooting would appear to be every bit as justified as hooking up a spring gun to do the same.

If, as we have suggested, none of the available solutions for avoiding the bootstrapping problem turns out to be compelling, we are left to consider whether we cannot after all make our peace with the idea that though it is impermissible generally to use lethal force in defense of property, it may be permissible to do so in the context of an overall justified plan involving the issuance of a legitimate deterrent threat. Similarly, though it would not be morally permissible to inflict a massive first strike on an enemy in wartime, it may be permissible to inflict such harm by way of retaliation for a deterrent threat justifiably issued that failed to deter. One must be careful here. It is not as though *any* deterrent threat and its follow-through can be justified by this bootstrapping argument. For sometimes the entire package—threat plus follow-through—is insufficiently justified by the interest we are trying to protect. Would it be legitimate, for example, to threaten to issue a nuclear strike to prevent someone from trespassing on one's lawn? Presumably not. The entire package—threat plus follow-through—must be justified in terms of the underlying interest being protected. Elsewhere, one of us has discussed this question in detail, and has tried to offer an account of how this kind of package justification might be thought to work.[59] It should at any rate improve the plausibility of the bootstrapping view to see that any such account will be subject to an internal proportionality requirement.[60]

In our view, the question explored in this section is part and parcel of a more general question in moral reasoning: Does embedding an impermissible act in a plan it is morally permissible to adopt change the moral status of the acts that make up the plan? In other words, does the moral permissibility of the plan translate into the moral

---

[58] Claire Finkelstein, *Acting on an Intention*, *in* REASON, INTENTION AND MORALITY (Gijs Van Donselaar & Bruno Verbeek eds., Ashgate Publishing,forthcoming 2008).

[59] *See* Claire Finkelstein, *Acting on an Intention*, *supra* note 58; *see also* Claire Finkelstein, *Constrained Maximization and Risk* (Oct. 20, 2005) (unpublished manuscript, on file with author).

[60] This is a point Alexander appears to miss, and he seems to suppose that given the existence of *notice*, any deterrent threat, and hence potential follow-through, can be made legitimate. In our view he misperceives the justifying element. It is not the notice itself that provides the justification for enhanced force, but rather what we might call the logic of the package deal.

permissibility of the individual acts called for by the plan? It is precisely the reverse of the question we addressed in connection with the *Actio Libera in Causa*, where we asked whether an otherwise *permissible* act (such as self-defense) can be rendered *impermissible* by the fact that it partakes in an overall plan it was impermissible for the agent to adopt. In the next section, we turn to a more specific comparison of the *Actio Libera in Causa* situation and the problem of deterrent threats, in the hope that their juxtaposition will serve to illuminate each with the insights gained from the other.

IV. CONTRIVED DEFENSES AND DETERRENT THREATS COMPARED

We began by noting that the problem of contrived defenses and that of deterrent threats are mirror images of one another: The former involves prima facie permissible acts embedded in an overall impermissible plan, while the latter involves prima facie impermissible acts embedded in an overall permissible plan. If we are correct in identifying this symmetry between the cases, one might expect the various available positions on each topic to admit of a corresponding position on the other. This is indeed what we find when we directly compare the various positions we have explored. Let us begin by attempting to transpose the analysis we just gave of the problem of deterrent threats into the *Actio Libera in Causa* situation. We will then consider the reverse transposition, namely the implications of the various positions on the *Actio Libera in Causa* for available positions on deterrent threats.

First, notice that in both cases we have two stages, which together form the entire plan. Suppose we begin by correlating the threatening stage in the deterrent threat situation with the contrivance stage in the *Actio Libera* situation. The threat execution stage then corresponds to the stage at which the defendant performs the prima facie criminal act in the *Actio Libera* situation. In both cases there is an overall plan that ties the first and the second stage together.[61] Now recall our first premise in the deterrent threat argument. There we saw that someone might wish to deny the permissibility of threatening to do more than it is actually permissible to do. Someone who took this position would reason, by backwards induction, about which threats it is permissible to issue, based on those actions it is permissible to perform at the second stage. The same kind of backwards induction would lead to a clear position in the *Actio Libera* situation. If we endorse the Backward Induction View in this case, we would be led to say that since killing with a justification or excuse is *permissible* (i.e., non-culpable), then the contrived plan, or the intention, to kill in that state would also be permissible.

---

[61] There is an asymmetry in the relationship between the first and the second stages, however, in the two contexts. The agent hopes *not* to reach the second stage in the deterrent threat situation, whereas his plan presupposes that he will reach the second stage in the *Actio Libera* situation. But this is just to say that his intent regarding the second stage is conditional in the deterrent threat situation and unconditional in the *Actio Libera* case, and it does not seem to make comparisons less fruitful across the two types of situations.

Take, for example, the defendant who contrives to kill someone in self-defense. Or consider the defendant who ensures that he is in a state of such irresponsibility that he cannot help himself from killing his victim. These defendants are intending to do something which, considered in and of itself, is perfectly permissible: killing someone in self-defense, or killing someone in a state of irresponsibility. So it looks as though the first of the positions we considered above, which denies the premise that it is permissible to threaten more than it is permissible to do, would translate into a denial that there is any problem about contrived defenses. The defendant still retains the benefit of any defense he contrives, and the larger plan invalidates nothing.

Now consider the mirror image of the position that says, on the contrary, it is *permissible* to threaten more than it is permissible to do. In the *Actio Libera* kind of case, this would translate into an argument that despite the permissibility of killing in self-defense, it is in fact impermissible to contrive to do that which it is permissible to do. The moral character of the act itself does not determine the permissibility of the intention leading to that act. Interestingly enough, many people who would endorse liability in the *Actio Libera* cases would also want to endorse liability in the case of deterrent threats and deny the permissibility of issuing such threats. A more natural position for someone inclined to favor liability in the case of contrived defenses, however, is to affirm the permissibility of deterrent threats.

Let us now turn to the second question we considered in the previous section, namely whether it is permissible to use an automatic retaliation device to make a deterrent threat more credible. The corresponding case in the *Actio Libera* context would be one in which acting on the contrivance was no longer under the defendant's control. Thus suppose the defendant has himself hypnotized to kill his wife. The operation of the hypnotic state is like the spring gun, in that the defendant proceeds to execute the plan on automatic pilot. Like the decision to set up the spring gun in the first place, the only voluntary actions were performed earlier, at the moment of contrivance, and the defendant was thereafter the proverbial "bullet in flight." The position of those who think that if manually defending property with deadly force is impermissible, then the use of spring guns is also impermissible, would translate into the position that if killing under hypnosis is permissible in the sense of being *excused*, then arranging for oneself to kill under hypnosis is also permissible. The automaticity of the mechanism by which one performs a justified action appears to be a more emphatic way of contriving or intending to perform a justified action.[62] The person who therefore thinks that one does not lose the defense if one contrives it should not be put off by the use of an automatic plan-execution device, such as hypnosis. The defendant who arranges for himself to automatically perform a permissible action would still be able to claim the benefit of the excuse or justification.

---

[62] The point is a bit trickier for excused than for justified criminal behavior, as it is hard to consider killing in an irresponsible state as *permissible* in the way we would regard killing in self-defense as permissible. But this is a wrinkle we will overlook here, as it seems likely some structural adjustments could be made in the argument to account for it.

Finally, consider the third step in the argument in favor of following through on permissible threats in the deterrent threat case. That stage, to recall, maintained that if it is permissible to back up a deterrent threat with an automatic retaliation device like a spring gun, it must be permissible to do the same manually, without the automatic device. This was what we called the bootstrapping argument. What is the bootstrapping argument in the *Actio Libera* situation? It would be the suggestion that if it is impermissible to contrive to kill someone "automatically," i.e., through a device like hypnosis, then the actual killing performed under hypnosis must be considered impermissible as well. In other words, we can hold the defendant liable not only on the basis of what he does at the planning stage, but on the basis of what he does pursuant to that plan. On this view, we would have bootstrapped our way into an argument against the permissibility of certain kinds of excused or justified actions, despite the fact that the defendant's conduct as such merits a defense. On this view, we must examine the larger circumstances in which a defendant acts in order to know whether he truly merits a defense it would otherwise appear he can claim.

Thus the bootstrapping logic that looked so questionable in the context of deterrent threats now seems perfectly palatable in the context of the *Actio Libera in Causa*. That logic suggests that we must assess plans—and component parts of those plans—in the context of the overall moral character of the entire package. As we have seen, the legitimacy of a particular deterrent threat must be assessed in the context of the overall justifiability of the plan to threaten, and eventually to follow through, in order to deter someone else's wrongdoing. And just as we saw there must be a proportionality restriction on the kinds of threats we can justify with bootstrapping arguments, so there must be restrictions on which contrived defenses we can invalidate with this method. Contriving to steal a loaf of bread by starving oneself until one is in a life threatening condition may not warrant invalidating the defense for the sake of recognizing the impermissibility of the contrivance, just as threatening the use of nuclear weapons is not warranted to deter trespassers. But these substantive considerations would require much greater elaboration before the plan-based logic we have suggested would provide a clear method of analysis. Here we have sought only to establish the parallels between these two problems in the criminal law, and to suggest a way of reasoning about both types of situations that appears to produce plausible conclusions.

So far we have examined what the different approaches one might take to the deterrent threat problem would imply for the problem of contrived defenses. Let us now consider how our analysis of the deterrent threat problem helps us to deepen our understanding of the different theories scholars have proposed to deal with contrived defenses: the traditional approach, the perpetration-by-means approach, and Joachim Hruschka's *Ausnahmemodell*, or secondary duty theory.

To simplify things, let us simply consider the two most extreme approaches people have taken to the deterrent threat problem. At one end, there is the approach that says that since it is immoral to kill someone to defend one's property, it is also immoral to intend to do so, to threaten to do so, or to set up a mechanical device that would do so. Since this approach has us reason from the immorality of the final act to

the immorality of all earlier behavior that produces it, we noted that this approach is sometimes called the Backward Induction View. At the other extreme was the approach that said that since it is permissible to threaten more force than one is ordinarily allowed to employ, it is also permissible to set up a mechanical device that would kill the thief, or even to kill the thief oneself, provided one does so as part of a general defense strategy. We called this approach, which has us reason from the morality of the overall defensive strategy to the morality of all of its component actions, the Bootstrapping Argument.

If we look at the traditional approach in light of the Backward Induction View and the Bootstrapping Argument we might notice some interesting parallels. The first is that the traditional approach is very similar to the Backward Induction View and very different from the Bootstrapping Argument. Like the Backward Induction View, it has us reason from the immorality of the individual actions of the defendant to the overall morality of his actions. The traditional approach says that the defendant is doing nothing wrong when he defends himself at the second stage, but that he is doing something wrong when he sets things up that way at the first stage, since at that first stage he cannot claim the benefit of self-defense. That is precisely the sort of act-by-act logic exhibited by the Backward Induction View.

There is something else, however, which the Backward Induction View can teach us about the traditional approach. Upon close examination of the parallel between the two, it becomes apparent that the way the traditional approach uses the act-by-act logic is defective. Instead of saying that at T_1 the defendant is causing death at T_2 and is doing so at a time when he cannot claim self-defense, a different description of the situation would be more consistent with the Backward Induction View. For that view suggests we focus on what the defendant is doing at the final stage, T_2,which in this case is killing in self-defense, and then to infer from that the moral status of his behavior at the initial stage, T_1. On this way of thinking, the Backward Induction View actually suggests that we should say that the defendant at T_1 is causing death at T_2 *by way of self-defense*. If we put the matter this way, it would seem as though the defendant is doing nothing wrong, since he is causing himself to do something perfectly legitimate, namely kill in self-defense. What the Backward Induction View reveals is that the traditional approach does not do what its adherents claim it does: it does not in fact help to rationalize our intuition that a defendant should not be able to make a self-defense argument if the defense was contrived. Quite the contrary, it suggests that he should be able to do so.

If we were to reconsider the perpetration-by-means approach in light of the foregoing discussion, we would find something analogous: the perpetration-by-means approach, like the traditional approach, is actually quite close to the Backward Induction View, as it too relies on a kind of act-by-act logic according to which we must infer the morality of a course of action from the morality of its individual parts. And one would finally find that the perpetration-by-means approach, like the traditional approach, has generally been misunderstood. Correctly applied, this view also suggests that the creator of contrived defenses should be acquitted rather than convicted.

Things stand very differently with Hruschka's secondary duty theory. On close examination, Hruschka's theory turns out to be quite similar to the Bootstrapping Argument and very different from the Backward Induction View. That is because fundamentally Hruschka's approach has us evaluate the defendant's course of conduct in its totality, and then assess the individual components of that course of action accordingly. Hruschka has us look at the course of conduct as a whole, and if we find that we disapprove, to recognize a secondary duty not to create situations that would require the defendant to perform the illegal, but ostensibly justified or excused, action. The Bootstrapping Argument would in parallel fashion have us look at the entire course of conduct of the designer of the deterrent threat arrangement, and if we find that we approve of it, to approve as well of its individual components, including the decision to use deadly force to kill a thief at the second stage. In this way we can see that Hruschka's approach is actually the only one that can satisfactorily rationalize the intuition of those who want to deny the contriver his defense.

That does not prove that the secondary duty theory is right, but it puts a high price on its rejection. For if we reject it, the logic of such cases will push us to approve of the use of contrived defenses, which many find unintuitive, and to reject deterrent arrangements which many find more palatable.

V. CONCLUSION

We have presented a long and complex argument designed to demonstrate the parallel between the problem of contrived defenses and that of deterrent threats. In our view, these two situations display a surprising symmetry once one notices that the agent in each case causes himself to perform an action whose moral character is at odds with the moral character of the overall plan from which he acts. Realizing this symmetry one acquires a potent tool for discriminating among different solutions to each of the problems. We are inclined to conclude that the resolution that proves able to deal most consistently and least counterintuitively with both problems is something along the lines of Hruschka's secondary duty solution to the *Actio Libera* problem, on the one hand, and the Boostrapping solution to the deterrent threat problem, on the other. This also suggests that plan-based approaches to problems of rational and ethical choice should be taken seriously, as it appears that many of our intuitive responses to ethical dilemmas can be best rationalized once we adopt a plan-based perspective.