THREATS AND CRIMINAL DETERRENCE IN SEVERAL DIMENSIONS

*Saul Levmore**
*Ariel Porat***

This Article's claim is that deterrence in criminal law is composed of four oft-competing goals that ought to be taken into account when calibrating sanctions. Consider an offender, such as an armed robber or kidnapper, who threatens to commit a yet more serious crime unless the victim surrenders. The first goal, or dimension, of deterrence is to deter the threat, or crime in general, and this suggests a high sanction for the armed robbery or other threatening crime. A second goal is to deter escalation, in this case the execution of the threat once issued. This goal normally suggests a large gap between the sanctions for lesser and greater crimes, in order to discourage escalation. Given a natural, or practical, ceiling to criminal penalties, the suggested gap often implies a relatively low sanction for the lesser crime, or threat, and some sacrifice of direct deterrence.

Two other goals, or dimensions, of deterrence have been unrecognized by academic commentators and legislatures. By extending the analysis to include behavior by victims, it becomes apparent that law ought to make the offender's threat incredible to victims. If the threat can be weakened, victims will be less inclined to submit, and then offenders will have less reason to threaten in the first place. This third dimension suggests a large gap between the sanctions for the lesser (threat) and greater (execution) crimes; if the gap is small, victims will recognize that the offender has little to lose from continuing on and executing his threat, and these victims will submit and make the criminal's threat profitable, thereby encouraging yet more criminal activity.

A fourth, equally novel, dimension of deterrence is the ability of law to make incredible the implicit promise by a threatener not to escalate if the victim submits. When this implicit promise is incredible, victims will again tend not to submit to the threat, and offenders will,

* Saul Levmore is the William B. Graham Distinguished Service Professor, University of Chicago Law School.

** Ariel Porat is the Alain Poher Professor of Law, Tel Aviv University, and the Fischel-Neil Distinguished Visiting Professor, University of Chicago Law School.

in turn, issue fewer threats. To achieve this goal, the sanction for repeating a threat following submission by the victim should be low; victims will be less apt to submit to a threat if they realize that submission will often bring on another crime, or threat. The idea—which should be applied with caution—is familiar from blackmail; a victim who thinks the blackmailer will return for further payments will not pay the first time, and the blackmailer, in turn, will be discouraged.

Optimal deterrence requires that criminal law should be attentive to the four dimensions described above. This Article presents a model for how law can better accomplish its deterrent function.

TABLE OF CONTENTS

I. INTRODUCTION

Ancient law punished only completed crimes, but law has evolved to sanction many attempts and threats.[1] Moreover, many completed crimes are platforms for escalation to more serious crimes, so that many

1. *See* Thomas Bittner, *Punishment for Criminal Attempts: A Legal Perspective on the Problem of Moral Luck*, 38 CANADIAN J. PHIL. 51 (2008). The punishment of attempts is easy to understand. When someone is apprehended just before committing a robbery, murder, or other serious crime, it is almost impossible to imagine a developed legal system, or professional police force, that would be satisfied with preempting the crime and then simply warning the nascent criminal. *See also* LEO KATZ, BAD ACTS AND GUILTY MINDS: CONUNDRUMS OF THE CRIMINAL LAW 276–300 (1987) (discussing attempts).

crimes constitute implicit or explicit threats.[2] Criminal activity often begins with a threat, or other low-level crime, because an offender prospers if he can acquire something of value with mere words, the wave of a weapon, or a low-cost wrong. These threats, as we will call them, are often cheap to inflict but terrifying to the victim. Modern law responds to this reality by criminalizing "specific" threats as well as gateway crimes, but this Article argues that the strategy is often misguided. Criminal penalties surely deter some threats, including low-level crimes that serve as implicit threats, and may reduce crime by removing convicted threateners from the population, but, once a threat is made, the penalty makes it *more* likely that an offender who threatens will carry out his threat.[3] The offender may be insufficiently deterred because the additional penalty for going beyond the threat is smaller than it would be if the entire deterrent were marshaled against the second step, or actual execution. In most areas of criminal law, it is reckless to improve marginal deterrence by simply lowering the penalty for the less-severe crime because this underdeters the less-severe crime. In the case of pure threats, there is more room to reduce the penalty for the threat, or first step, but to see why this is so requires an understanding of the credibility of threats.

These familiar deterrence goals are incomplete because they are based on a fairly static analysis of the threatener's perspective alone. The value of the threat to the offender depends in large part on the victim's perceptions and likely response. If the victim finds the threat of escalation credible, and submits to the offender, the offender is better off, and wrongdoers are encouraged to make such threats again. A novel argument in this Article, referred to as "primary credibility," is that by lowering sanctions for threats, law could make threats less credible and, therefore, less valuable to offenders. The counterintuitive idea is that when

2. The marginal deterrence factor is less important for attempts because, at least in principle, law deters completed crimes by deterring their attempt. There is no time lag between the two, as there is for a threat and its execution. It is easy to see why an offender apprehended one moment before pulling a trigger should be punished and, if this were not the case, police would face impossibly difficult problems as they sought both to prevent harms and apprehend those likely to commit them. It is because a threat can be an alternative, or a cheap substitute, for a completed crime that it is of special interest. *See* Steven Shavell, *A Note on Marginal Deterrence*, 12 INT'L REV. L. & ECON. 345, 346 (1992).

3. For a pathbreaking and excellent technical discussion of this intuition—and why it is not always correct once one takes costs and substitutions into account—see David Friedman & William Sjostrom, *Hanged for a Sheep—The Economics of Marginal Deterrence*, 22 J. LEGAL STUD. 345, 345–46 (1993) ("In such a situation, one of the considerations in setting punishments is the risk that a high punishment for one crime may shift the offender to committing a different, and perhaps a worse, one."). *See also* Shavell, *supra* note 2 (showing that where criminals can substitute between two acts, optimal sanction for one might well be below the harm caused by the other to influence choice of acts); George J. Stigler, *The Optimum Enforcement of Laws*, 78 J. POL. ECON. 526, 528 (1970) ("Marginal costs are necessary to marginal deterrence."). The basic marginal deterrence idea has been recognized for 250 years or even 337 years in the Anglo tradition. *See* Friedman & Sjostrom, *supra*, at 345–46 nn.1, 5. For the best-known early statement, see JEREMY BENTHAM, THEORY OF LEGISLATION 201 (2d ed. 1871) ("Where two offenses are in conjunction, the greater offense ought to be subjected to severer punishment, in order that the delinquent may have a motive to stop at the lesser."). Skepticism about marginal deterrence theory, and discussion of empirical evidence about general and specific deterrence, is best associated with another equally pathbreaking work, FRANKLIN E. ZIMRING & GORDON J. HAWKINS, DETERRENCE: THE LEGAL THREAT IN CRIME CONTROL (1973).

the sanction attached to a threat is high, the sanction for executing the threat becomes lower. When that latter sanction is lower, the threat becomes more credible and victims are more likely to submit. Submission increases the offenders' profit and criminal activity. By lowering the sanction for threats and focusing law's deterrence power on their execution, law can reduce the profitability and volume of criminal activity. To be sure, the law might distinguish situations in which it is plausible that victims have the time and capacity to make considered responses to threats from situations in which they do not.

Our attention to credibility is a means of incorporating the victim's perspective and likely response to a threat. It enables a dynamic analysis in which the threatener's behavior depends, in part, on the victim's expected response, which depends, of course, on the perceived threat. In turn, the analysis reveals a fourth dimension, or goal, of deterrence, that has also gone unnoticed in the legal literature. We refer to it as "secondary credibility." The effect of a threat depends in large part on the credibility of the offender's implicit promise not to execute the threat, or not to re-threaten, if the victim submits to the threat in the first place. Counterintuitively, if the sanction for re-threatening is low, victims will realize that if they surrender to the first threat, another may well follow. This can discourage submission in the first place, especially when victims have time to plan ahead or think through their responses to criminal wrongdoing. In the long run, reduced submission makes criminal activity less profitable.

The new dimensions introduced in this Article can be distinguished from the more familiar marginal deterrence arguments. The latter aim to deter the escalation of a crime; the armed robber, for example, is hopefully deterred from killing witnesses. In contrast, the credibility perspective adopts an earlier vantage point and seeks to deter criminal activity at the outset.

These four dimensions of criminal deterrence—direct deterrence, marginal deterrence of escalation, and then primary- and secondary-credibility effects—can be illustrated by considering a threatener, T, who targets a victim, V. T threatens force unless V gives up something valuable. Assume that the sanction for using force is 20, and the sanction for making the threat is 8, or alternatively, 1. When the penalty attached to the threat is 8, the sanction for its execution, which is to say the actual use of force, is only 12. That might be too low to deter T from executing the threat he has issued.[4] If the sanction attached to the threat alone were 1, however, T might not execute on his threat because of the greater *marginal deterrence* provided by the sanction of 19, applied to an offender who uses force. The direct-deterrence dimension likewise considers the threatener's perspective. If the penalty attached to the threat is lowered in order to increase marginal deterrence, there is the likelihood that

4. The actual threat of the sanction to the criminal is discounted, of course, by the probability of apprehension. *See infra* note 34.

offenders will be under-deterred and will issue more threats. This direct, rather than marginal, deterrence consideration incorporates a *substitution effect* because offenders can be expected to gravitate toward, or substitute, threat-making as its sanction is reduced.[5] This substitution is insignificant if threats themselves are harmless, but it is plausible that would-be offenders underestimate the injuries they will cause if they wave guns in the air. And there are other reasons to be concerned about threats; they impose anxiety costs, and some victims will submit to them without pushing the threatener to decide whether or not to execute.

The third dimension is *primary credibility*, and its novelty derives from incorporating the victim's perspective within deterrence theory. If T is often undeterred by the marginal penalty of 12 rather than 19, then V will find T's threat more credible when the sanction for the threat is 8 rather than 1. The less T is deterred from violently completing or escalating the crime, the more V finds the threat of force credible. V will be more likely to submit when threats are sanctioned and, correspondingly, a low sanction for the threat might therefore *reduce* credibility and submission. The primary-credibility effect refers to the feedback on T. As decreased marginal deterrence for escalation makes T's threat more credible to V, T is encouraged to threaten because V is more likely to submit. Remarkably, when law penalizes threats severely, it can make threat-making—which is to say,most crime—more profitable.

Finally, there is the dimension of *secondary credibility*, which is most apparent in cases like blackmail, where repeat threats haunt a victim who has time to consider the best response to a threat. Imagine a legal system that punished only the first instance of successful blackmail and provided no further penalty for a blackmailer who, after receiving payment from a submissive victim, returned for more. Knowing this at the outset, V is more likely to stand up to the blackmailer and refuse to pay and, in turn, blackmail would become less profitable.

It is apparent that law's treatment of threats should depend on a dynamic analysis that takes into account both the offender's and victim's perspectives. To best structure criminal penalties, it is necessary to understand exactly how threats work and what makes them more or less credible. Much as the marginal penalty expected by an offender affects the victim's decision to submit to a threat, so too the credibility of the threat as perceived by the victim affects the offender's threat-making. The less credible a threat, the less likely a victim is to submit, and the less profitable the criminal activity. Part II, therefore, begins with an exploration of the ingredients of credibility. The discussion continues to distinguish between primary and secondary credibility. The former refers to the likelihood that the threat will be carried out if the victim fails to submit. It depends on execution costs, as the victim must believe that the threatener will find it worthwhile to carry out his threat if the victim fails

5. The substitution label avoids the need to call it an underdeterrence, or inadequate direct deterrence, problem. It also avoids confusing marginal and direct (activity-level) deterrence.

to submit. Secondary credibility refers to the reliability of the threatener's claim that submission will indeed preclude execution or escalation. The discussion concludes by challenging the conventional wisdom that repeat play guarantees the credibility of threats.

Part III sets out the four dimensions of criminal deterrence. It begins by developing the marginal deterrence and substitution (direct deterrence) effects from the offender's perspective. The discussion suggests that, in its different treatment of specific threats, as opposed to more general threats, law may already reflect a sophisticated understanding of the interaction between marginal deterrence and substitution.[6] This distinction offers a window into the argument that we make about the benefit of allowing threats to go lightly punished. The analysis then introduces victim behavior and explores the interaction between criminal sanctions and the credibility of threats. The discussion suggests a revamping of criminal penalties.

Part IV extends the analysis to include the risk that offenders can sink costs to enhance credibility. When threats are involved, these sunk costs enhance the credibility of the threat exactly as criminal penalties do when attached to mere threats because both reduce the marginal deterrence to escalation. The strategy is most plausible where the first step involves a significant fraction of the overall costs. In turn, it can be a good strategy to deter the offender at this first step, even though marginal deterrence with respect to the second, more serious, step is compromised. The discussion explores the limits of this argument.

II. ASSESSING THREATS AND THE DANGER OF ESCALATION

A. Prerequisites for Credibility

Targets will not respond to threats in a way that makes the threatener better off unless the latter's threats are credible and plausibly injurious. "Give me your wallet or I will shut down your university for a day," is neither and is more likely to elicit puzzlement than fright or anyone's wallet. Threats are most credible when the cost of execution is low compared to what the threatener is likely to gain.[7] An extortionary arsonist needs but a can of gasoline and a few matches to make good on his threat; if his threat to burn a shop owner's place of business is credible, he might be able to use the same tools repeatedly against many victims, and he might not even need to use the tools at all.

A threat can be contrasted to its opposite number, a contractual promise, which pledges a benefit rather than a harm. In the absence of

6. *Cf.* Pedro Celis, *When Is a Youtube Video a "True Threat"?*, 9 WASH. J. L. TECH. & ARTS 227, 229 (2014) ("Requiring subjective intent reduces the potential chilling effect of § 875(c) by ensuring that only threats directed at specific individuals or groups are subject to liability.").

7. Steven Shavell, *An Economic Analysis of Threats and Their Illegality: Blackmail, Extortion, and Robbery*, 141 U. PA. L. REV. 1877, 1878 (1993); *see also* Guiseppe Dari-Mattiacci & Gerrit De Geest, *Carrots, Sticks and the Multiplication Effect*, 26 J. L. ECON. & ORG. 365, 365–66 (2009).

law, a threat of harm is often *more* credible than a promise of a benefit because the promisor's execution costs are usually high.[8] For example, when an employer promises a promotion and raise, the employer must actually expend money to fulfill the promise and, of course, the promisee must generally work hard to earn this money. In contrast, as noted, an offender can threaten by waving a stick or other weapon. Understandably, law intervenes to raise the offender's execution costs; the extortionist can wave a stick or fill a can of gasoline, but criminal law causes him also to anticipate some probability of apprehension and criminal penalty. In turn, shopkeepers can forego some defensive measures because they count on law to deter threatening offenders.

Assume that a threatener, T, makes a threat and has the power to execute it, and that T's target, or victim, V, knows this to be the case. Assume further that V will find it cheaper to submit to T than to suffer the injury that T can bring about if he executes. When these conditions are met we say that the threat has *primary credibility*. Of course, V may submit and then find that T executes anyway. Prior to submission, V must therefore be confident that submission will not be followed by execution and that T will not return at a later date to renew the threat. Thus, one reason blackmail is not more common is that the target fears that the blackmailer will demand yet more money, once the blackmailer learns that the target will pay for the photograph (or other evidence regarding a point of vulnerability) in question.[9] We call these matters of *secondary credibility*. Note for now that when V contemplates submission, V must estimate both primary and secondary credibility—the likelihood that T will carry through on his threat (primary credibility) and also that T will not execute despite submission by V or threaten again (secondary credibility). T, in turn, wants to be perceived as secondarily credible in order to raise the probability that V will submit. The conventional means for T to establish credibility is to have a history of reliability through repeat play, a topic taken up below.[10]

However credible T's threat, V may, at times, be able to neutralize the threat rather than submit to it. For example, V might respond to T's threat of "your money or your life" by taking out her wallet and quickly destroying its contents. Few criminal threats offer V this defensive op-

8. Saul Levmore & Ariel Porat, *Rethinking Threats*, (Coase-Sandor Inst. L. & Econ. U. Chi. L. Sch., Working Paper No. 721, 2015), http://chicagounbound.uchicago.edu/law_and_economics/738/. *See* Dari-Mattiacci & De Geest, *supra* note 7, at 367.

9. *See* Ronald H. Coase, *The 1987 McCorkle Lecture: Blackmail*, 74 VA. L. REV. 655, 671 (1988) (arguing that blackmail does nothing more than transfer wealth and that because there is no reason to think that it is a transfer to a higher valuing user, and because resources are wasted in procuring and hiding information, it is sensibly prohibited by law).

Blackmailers face problems with respect to both primary and secondary credibility. The first problem is one of correct pricing. If the blackmailer starts high and lowers his price in the event of non-submission, his threat seems weak because his primary credibility is weakened. And if he starts low and then raises the price in the event of apparent submission, he seems secondarily unreliable.

10. *Infra* Section II.C.

portunity, though such "scorched earth" strategies are discussed later in this Article.[11]

B. Execution Costs and Benefits

When T says, "Your money or your life," V knows that execution is more likely the more money T thinks V possesses, and the easier it is for T to grab that money without apprehension or injury to himself. Primary credibility is therefore most likely where there are low execution costs. And yet, secondary credibility is enhanced when V perceives that T's execution costs—following V's submission—are high. This apparent contradiction, or offset, is penetrated by considering T's varying execution costs and benefits before and after V submits to the threat. In particular, even when execution following submission is of low cost, if there is nothing else for T to gain after V submits to the threat, V may well find T's threat primarily and secondarily credible. We will return to this point when we discuss how criminal law affects credibility and how law might manage T's behavior following submission by V.[12]

The significance of execution costs is made plain by considering examples outside of the normal reach of criminal law, where execution costs and credibility are divorced from the probability of apprehension. If T says, "My group will boycott your business unless you do *X*," or, "I will go on a hunger strike until you do *Y*," V can assess the cost of execution to T and calculate that the threat is increasingly incredible as that cost rises. Setting aside the value of adding to T's reputation for reliability, the cost-benefit calculations are straightforward. If T threatens a twenty-day hunger strike unless V volunteers for one day at the local homeless shelter, V will doubt the threat's credibility, and will regard the threat as a signal of intense feelings, to be sure, but not much more.[13]

Matters can be further complicated when the *victim's* secondary credibility is in play. Just as V must worry that submission will not guarantee safety, there are situations where T must be concerned about V's reliability. Imagine that T threatens his employer: "I see you have discovered my embezzlements. If you report me to the police, I will burn down your warehouse." If V submits, and the two parties agree that V

11. *See infra* note 54.

12. *Infra* Section III.D.

13. Even where apprehension and criminal punishment are excluded, the parties can have difficulty assessing one another's costs and benefits and, therefore, credibility. Imagine a valued faculty member who threatens his law school's dean as follows: "Raise my pay by $50,000 or I will move to Law School Z." It is difficult for the dean to assess the threatener's cost and benefit of execution because of the obvious information asymmetry; only the threatener knows whether a move is costly or, in fact, much desired from the threatener's perspective. It is difficult if not impossible to make such a threat more credible without irrational self-destruction on one side or the other. Consider, for example: "My threat is real. If you do not submit, I will be gone, and to show you that I am not bluffing, if I do stay without getting a raise, I will owe the law school, $100,000." No dean would "accept" this promise, or seek to enforce it, because it encourages the very defection that the dean seeks to avoid (although in the long run it might reduce defection). It is apparent that the high cost of execution makes some threats less credible, but in virtually all cases it adds to secondary credibility.

will allow T to separate from the workplace quietly, there is the possibility that V will surprise T and have him arrested one year later. It is not easy for T to overcome this credibility problem, and in turn it is likely that such threats will be rare.[14]

C. Repeat Play

If T is a repeat player, then it seems elementary that T's threats are more credible because V will calculate that T has reason to execute in order to establish a reputation as a credible threatener. T is also more reliable, or secondarily credible, than nonrepeat or anonymous threateners. In the most straightforward case, a criminal enterprise can demand "protection" payments and garner an attractive, extortion-induced revenue stream with just the right level of threats and execution. Its primary credibility is enhanced because a victim knows that the threatener has more to gain from execution than does the typical mugger; future victims have no way of knowing whether a mugger who confronts them has previously resorted to violence, but a storekeeper asked to make protection payments by a criminal enterprise knows that violence by this threatener is an investment that will cause other storekeepers to submit and may also know of prior violence.

The organized criminal enterprise further enhances secondary credibility by muscling out competitors so that victims will know that, if they submit, their vulnerability will not make them prey to other offenders.[15] A terrorist enterprise that is organized around an ideology may be even more formidable. If its frontline soldiers are trained to welcome martyrdom, then they are known not to fear the serious criminal sanctions attached to execution, and this improves primary credibility. And if the organization controls a population, then it is secondarily credible because of repeat play and its ability to prevent copycat threats.[16] Indeed, a government that pays a terrorist group to release hostages must have reason to believe either that the group has the means to prevent repeat terror, or that the government can prevent repeat terror and is simply paying for the release of the hostages.[17]

14. It is tempting to say that submission suffers from the same credibility problem as threats, or indeed that V implicitly "threatens" to reverse any submission. Working backward, then, T will seek to make a demand, like handing over a wallet, that V cannot easily reverse. Similarly, it is in T's interest to make a threat that is both immediately and secondarily credible. V should submit, believing that submission will preclude execution. As will become apparent, this is part of the strategy of increasing marginal deterrence for executing a threat, by decriminalizing the threat itself.

15. *See* Thomas C. Schelling, *What Is the Business of Organized Crime?*, 20 J. PUB. L. 71, 73 (1971) (arguing that one characteristic of organized crime is that it allows no competition and exercises a monopoly); Shavell, *supra* note 7, at 1882 (stating that a threatening party may want to carry out a threat when its demand has been rejected for the purpose of establishing a reputation).

16. This repeat play is more significant than most because the organization's goal makes it a long-term player, without the end-period problem discussed presently in the text.

17. Harvey E. Lapan & Todd Sandler, *To Bargain or Not to Bargain: That Is the Question*, 78 AM. ECON. REV. 16 (1988) (arguing against the conventional wisdom that governments should not bargain with terrorists over hostages to avoid encouraging further hostage-taking). Criminal enterprises enhance credibility in other ways. They reduce the end-game problem by operating over long time

But the conventional wisdom, or intuition, about repeat play and credibility is overstated. For one thing, even where repeat play surely does enhance credibility, the matter is complicated both by the likelihood that V is also a repeat player, trying to build a reputation for non-submission, and the possibility that repeat play has an asymmetric impact on reputation. Furthermore, past reliability does not provide much information about the future, except that it is better than past unreliability. V does not know whether T is planning to make future threats or is in fact in the last phase of a sting of sorts. Consider an identifiable, serial, armed bank robber facing a victim who refuses to submit. If it is this repeat player's last hurrah, it is less profitable for him to shoot than it was earlier in his criminal career. Firing his weapon and maiming a teller or officer may allow the robber to steal a bank's cash, and thus obtain a direct benefit, but there is no need to invest in reputation, as this is his last robbery. In contrast, someone who reliably performs as required by contracts may be setting up a target for a sting. A contractual partner must always have her guard up in case this is the promisor's final act.[18]

In sum, repeat play hardly guarantees credibility. In the absence of repeat play, significant execution costs call credibility into question. Threat-making is thus a tough line of work because credibility is not easy to establish. But of course some wrongdoing begins with threats because credibility is within reach and because when submission is forthcoming, crime is a very profitable activity. The question in these settings is when law should focus its deterrent power on threats and when it should reserve this power to forestall the serious harms brought about by the execution of that which was threatened.

III. MARGINAL DETERRENCE, SUBSTITUTION, AND CREDIBILITY EFFECTS

This Part explores four dimensions that must be taken into account by legislatures when setting sanctions for criminal activities initiated by threats, including crimes that serve as threats of escalation to more serious crimes. The discussion begins with marginal deterrence and uses this familiar tool to explain a fundamental distinction in criminal law between specific and general threats. We argue that a plausible explanation for the decriminalization of general threats is the legislature's desire to reserve enough marginal deterrence for the execution of the threat. We then turn to the substitution effect—the risk of sacrificing direct deterrence if marginal deterrence for execution is enhanced—and explain the

periods; they lower execution costs through training and professionalization; they lower the expected cost of apprehension by investing in relationships with police and with legal-defense teams and by working to make any prison terms less painful for their members than they are to other convicted felons. They also enhance credibility through vigorous and well-advertised executions when targets fail to submit.

18. Repeat play might, however, show the target that the threatener has no fear of the police or is a "bad type." Still, the threat is less credible if the threatener is operating in his final period.

tradeoff between marginal deterrence and substitution. Finally, we introduce two new dimensions, primary and secondary credibility. When the sanction for threats is low, the primary effect is sufficiently powerful to overwhelm the substitution effect. Counterintuitively, the lower the sanction for threats, the less threats will be used by offenders. The analysis shows that when the sanction for escalation after submission, or for rethreatening, is low, threats become less credible.

A. General and Specific Threats

Criminal law has been unresponsive to, or dismissive of, general threats, or nonspecific threats made to a group of people. The typical requirements for a "criminal threat" include a victim in a state of reasonably sustained fear for his or her personal (or family member's) safety as the result of a threat that is specific and unequivocal.[19] The offender need not actually intend to execute the threat and, indeed, may not have the ability to carry it out.[20] Thus, we can expect a finding of criminal wrongdoing when someone points a gun at a stranger and announces, "Your money or your life," because the victim is singled out and has reason to fear bodily harm.[21] In contrast, consider one who threatens: "I cannot stand the noise you are all making; I hate you; if you do not shut up I'm going to come back and make you sorry." In normal circumstances, this threat is nonspecific both because it is aimed at a vaguely described group, so that no one has particular reason to fear for his or her own safety, and because it is equivocal in terms of timing and method. This lack of specificity is apt to make a criminal conviction impossible as a matter of law.[22] Without specificity, behavior may be regarded as wicked—but it is often not criminal until there is execution. At the very least,

19. *See* CAL. PENAL CODE § 422 (West 2016) ("[T]hat threat actually caused person threatened to be in sustained fear for his or her own safety or that of his or her immediate family; and that threatened person's fear was reasonable under the circumstances"); People v. Toledo, 8 P.3d 1051, 1055 (Cal. 2000).

20. *See Toledo*, 8 P.3d at 1051.

21. In Illinois, for example, the threat is referred to as an "intimidation." 720 ILL. COMP. STAT. 5/12-6(a) (2017) (including threats of physical harm to the victim, third parties, or property). Intimidation is a Class 3 felony, punishable by two to ten years of imprisonment. Once there is contact, or battery, the use of a firearm increases the penalty substantially. Battery is a Class A misdemeanor punishable by up to one year in prison. *Id.* § 12-3(b). Aggravated battery using a firearm occurs when T discharges a firearm and causes injury to V. *Id.* § 12-3.05(e)(1). This elevates the crime to a Class X felony (six to thirty years), where the penalty also varies with other factors, such as the identity of the victim.

Note that even if the threat described in the text rarely yields a profit to the wrongdoer, it is correspondingly inexpensive to impose. One weapon can be waved many times until a vulnerable target submits. The same is not true for a beneficial promise, which in some ways is the mirror image of a threat, and it is interesting that law has experimented with formalities and suits for dashed expectations primarily where promisors can be promiscuous, as with charitable gifts or expressions of intentions to marry. *See* Dari-Mattiacci & De Geest, *supra* note 7.

22. *See* Glenda K. Harnad et al., *Nature of Threat,* 19 CAL. JUR. 3D CRIMINAL LAW: MISCELLANEOUS OFFENSES § 186 (2017) ("Unequivocality, unconditionality, immediacy, and specificity are not absolutely mandated but must be sufficiently present in the threat and surrounding circumstances to convey gravity of purpose and immediate prospect of execution to the victim.").

some significant and active step is required to find that an attempted crime has been committed.

There are conventional explanations for this difference between specific and general threats. The exclusion of general threats from the reach of criminal law is, for one thing, a convenient way of saying that loose and angry talk should not normally be criminalized.[23] It may be sufficiently common for people to say things in fits of anger that a general understanding has developed that those who say ugly things when enraged, or under the influence of alcohol, do not have the intent required to trigger criminal penalties. Experience may have shown that very few of these general threats are harbingers of their execution, which is to say violence, whereas a much higher fraction of specific threats turns out to predict just what the threatener promised. At the same time, or perhaps equivalently, most listeners are accustomed to angry words that amount to nothing dangerous, or otherwise have no reason to be apprehensive when at the receiving end of such a threat. Rather than chilling speech or leaving an enormous pool of cases to prosecutorial discretion or juries' inclinations, law requires specificity. A companion argument is that, where general threats are concerned, it is more difficult to prove both intent and reasonable apprehension.[24] This latter explanation is more attractive because it is hard to see why speech directed at a specific individual, or simply containing sufficient details to qualify as a specific threat, does not enjoy as much protection as speech that is vaguer or is aimed at a dispersed group.

There is no reason to dispute this conventional understanding of general and specific threats, but, especially when the threat is of the general kind, because it was directed at a group of people, it is not obviously correct and it does not appear to reflect a social-welfare calculus. Within limits, the larger the group that is threatened, the greater the aggregate disutility, even if each person in the group has a small chance of becoming a victim. One who threatens a group usually creates more anxiety and likely stimulates more total precaution-taking than does one who threatens a specific person. For example, someone who threatens a shooting on a vaguely specified college campus may create sufficient anxiety to shut down a campus for a day. The anxiety and shutdown costs are great, even if no individual feels particularly threatened.[25] The requirement of reasonable apprehension seems sensible enough, but there is no reason to

23. *See* Virginia v. Black, 538 U.S. 343, 359 (2003) (explaining that the state can only punish threatening expression when the speaker "means to communicate a serious expression of an intent to commit an act of unlawful violence to a particular individual or group of individuals").

24. Put differently, if intent is important in criminal law, perhaps because it helps show that an injury was not accidental, then in turn an earlier, specific threat clarifies intent. Law and economics struggles with the intent requirement and often avoids the treatment of threats. *See, e.g.*, Richard A. Posner, *An Economic Theory of the Criminal Law*, 85 COLUM. L. REV. 1193, 1214–19, 1221–22 (1985) (rationalizing the treatment of attempts, struggling with intent, and avoiding threats).

25. *See, e.g.*, Jon Seidel, *University of Chicago Cancels Classes After FBI Warns of Gun Threat*, CHI. SUN TIMES (Nov. 29, 2015, 8:19 PM), http://chicago.suntimes.com/news/university-of-chicago-cancels-classes-after-fbi-warns-of-gun-threat.

expect that apprehension is more reasonable as the threat is addressed to a specific individual. In any event, not all general threats get a free pass. Law can and does criminalize group threats in piecemeal fashion. If, for example, there is a spate of school shootings, a legislature might make it a crime to threaten the use of a weapon or to bring a weapon onto school grounds.[26] Schoolchildren and their parents might have reason to grow anxious about threats made against an entire school, but the legislature can respond to their apprehensions about such a general threat, without directly removing the requirement of a specific threat from the law of threats.[27]

The distinction between general and specific threats, and the inclination to leave many of the former unsanctioned, may reflect an intuition about marginal deterrence. Marginal deterrence theory offers a means of drawing or understanding the line between specific and general threats that does not require any judgment about the value of one kind of speech or another, and that does not underestimate the harm imposed by a general threat.

B. Specific Threats and Marginal Deterrence Theory

The offender who points a gun at a pedestrian and says, "Your wallet or your life," presents law with the archetypal question of how much punishment to attach to the threat and then how much to leave for a completed crime, in this case a shooting. An important assumption here is that there is some ceiling to the penalty, or binding constraint on enforcement, so that the punishment for the threat cannot simply be x, for the shooting y, and for both $x+y$, as there would then be no reduction in marginal deterrence as x is increased.[28] Assume for the moment that law

26. Without such a statute, the remedies are limited. *See, e.g.*, Bolden v. Chartiers Valley Sch. Dist., 869 A.2d 1134, 1134 (Pa. Commw. Ct. 2005) (finding that employee's bringing loaded gun onto school property constituted neglect of duty, and the lying and hindering a lawful investigation initiated by the superintendent gave the district authority to take disciplinary action against employee).

27. As law criminalizes various general threats, albeit by assigning these threats to new categories like weapons violations or online crimes, the distinction between specific and general threats becomes formalistic. On the other hand, these legislative moves against various general threats mirror or confirm an analogy between general and specific threats, on the one hand, and tort actions and public nuisance, on the other. In the tort context, individual plaintiffs have standing to complain of a private nuisance, as in the case of a polluting factory next door, but where a public, larger-scale, nuisance is concerned—though it might be more harmful—the resolution of conflicting uses is left to the legislature. *See* Keith N. Hylton, *The Economics of Public Nuisance Law and the New Enforcement Actions*, 18 SUP. CT. ECON. REV. 43, 43–44 (2010) (discussing tradeoff between public and private enforcement). One rationale is that individual plaintiffs might have the wrong incentives when it comes to bringing or settling such claims. Similarly, in the criminal law, the more general the threat, the more it is poorly assessed by an individual or by a jury's assessment of the reasonableness of various individuals' fears. In turn, these are often settings where the legal system expects police or other authorities to keep the peace, dissipate any significant threats, and then proceed with a specialized criminal statute aimed at reckless endangerment, weapons violations, terrorism, or other categories that serve to circumvent the specificity requirement of the statute addressing criminal threats.

28. We assume away the likelihood of decreasing marginal disutility of punishment. In other words, y might provide more marginal deterrence when it is unattached to x because the offender fears the first units of punishment much more than later ones. We set this aside not only because potential

can solve this problem by simply reaching a reasonable balance between the threat and the completed shooting. It attaches a more modest punishment to the criminal threat, than that suggested by the proportionality principle alone, to leave space for a large enough marginal deterrent on the (threatened) shooting. Marginal deterrence theory is easily incorporated into law, and there is certainly no reason to assume that it is ignored by lawmakers. It is easy to imagine, for example, that the armed robber who threatens a pedestrian with his words and a drawn weapon can expect a three-year prison term if apprehended and convicted; a five-year sentence if he also steals the victim's wallet and runs off; a ten-year term if he shoots and maims the victim; and a thirty-year term if he kills the victim. Other steps, and associated numbers, can be inserted all the way up.

To be sure, a marginal deterrence theorist might say that three (years) is too big a number for the first of these wrongs because more (than two additional years') deterrence should have been reserved to discourage the taking of the wallet; but any such argument about the increments is empirical, difficult to resolve in the face of heterogeneous criminals who can change behavior over time, and perhaps wrong.[29] The discussion here assumes, optimistically and for the sake of exposition, that the numbers provided by existing law already reflect careful thinking about marginal deterrence.

Consider now the case of a general threat. Suppose an offender says to a group of partygoers: "Quiet down, or I am going to come back and make you sorry." A useful and novel way to think about why this general threat is not criminalized is that, first, the general threat need not cause anxiety or any harm on its own, and therefore deterring the threat as such is not so important, and second, because the details of any subsequent step are unknown and addressed to many people, law reserves its deterrents. The threatener has not said what he will do next, or to whom, and he has by assumption caused no physical or personal injury by the time he is detained. He may have become sober and calm, or he may have returned to find that the partygoers dispersed because of his threat, called the police, or readied themselves for his violent return. In the absence of good information about these later steps, it is perhaps clever to attach no penalty to the first step, the initial broad threat, to reserve deterrence for the subsequent steps, including attempted crimes. In this way, the distinction between specific and general threats can be understood through marginal deterrence theory.[30]

offenders are likely to be heterogeneous, but also because the shape of the disutility function matters little to our argument.

29. At the risk of repetition, see Friedman & Sjostrom, *supra* note 3, for special cases and assumptions where the marginal deterrence (quick) intuition is called into question.

30. It is tempting to add in the idea that this threatener of noisy partygoers is not credible, so that he imposes little anxiety and is, in fact, unlikely to pose much of a risk. One reason to doubt credibility is that the threatener has announced his plan, though not with specificity, and he can expect that defensive measures will be in place when he returns. But this is even more so if the threat is specific, as in, "Quiet down or I will return in two hours and kill the bartender." A specific threat of this kind is

C. Marginal Deterrence Followed by a Substitution Effect

How should theory influence the structure of criminal penalties? Assume that a starting point in the design of criminal law and penalties is the proportionality principle: the more severe the crime, the greater the penalty.[31] Assessments of severity might derive from straightforward deterrence calculations, retribution considerations, incarceration costs, or political pressures. This design principle yields a baseline, or schedule, of crimes and penalties. Imagine that along this schedule, theft from a person is assigned a penalty of one-year imprisonment; armed robbery, perhaps included in the category of aggravated assault, is assigned ten years; and armed robbery that includes a shooting in which someone is injured, sometimes known as aggravated battery, draws a penalty of twenty years.[32] For quick reference, we might summarize this part of the schedule as TH-1/AA-10/AB-20, where AB-20, for example, means that aggravated battery is associated with a twenty-year penalty.[33]

Marginal deterrence theory suggests that the legislature think carefully about the differential, or gap, between two crimes, wherever the offender is likely to be in a position to escalate from one crime to another. It might suggest that if the proportionality principle alone produces a ten-year gap between the two serious felonies on this schedule, it would be advisable to widen the gap in order to discourage the criminal from shooting his victim, as he might be tempted to do in order to gain property or eliminate a witness so as to escape apprehension. The legislature can do this by increasing the penalty on the more serious crime, producing a schedule such as TH-1/AA-10/AB-25, but this strategy quickly runs

more likely to trigger defensive measures and, in turn, execution seems less likely if the offender is rational. If the threatener warns the partygoers, "Get quiet right now or I will make you all sorry," the threat is incredible either because the threatener is outnumbered or because if he threatens to return with a weapon, there is time to mount a defense or call the police.

31. JEREMY BENTHAM, AN INTRODUCTION TO THE PRINCIPLES OF MORALS AND LEGISLATION 181 (1789) ("The greater the mischief of the offense, the greater is the expense, which it may be worth while to be at, in the way of punishment."); RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 258–74 (9th ed. 2014) (discussing the elements of an efficient criminal law and noting that much follows from the infeasibility of setting punishments high enough to achieve 100% deterrence).

32. The penalties are usually given in the form of ranges, so that aggravated battery with a firearm in Illinois, for instance, is a Class X felony, assigned six to thirty years. 720 ILL. COMP. STAT. 5/12-4.2(a)(1)–b) (2017). The same is true for theft, with penalties usually rising according to the value taken as well as whether the theft was from another person (as opposed to a store shelf, for example). In Illinois, the range is from a Class A misdemeanor to a Class 3 felony, and could generate a fine or a prison term. The mean penalty is probably between one and two years if the theft is from a person. *Id.* § 5/16-1(b)(1). The crimes and penalties used in the text are thus illustrative and show what the schedule would look like *before* further consideration of marginal deterrence, substitution, and the credibility effect described presently.

33. TH refers, of course, to theft from a person, and AA to aggravated assault, here the armed robbery with no discharge of the weapon. For simplicity, the example continues to assume that aggravated battery produces a sanction that is not added to the sanction for the aggravated assault, which will usually precede the battery. In Section II.B the two sanctions were not summed because of the ceiling on the system's ability to penalize. There is also a doctrinal reason for not summing the sanctions; the criminal charge of aggravated battery will include the lesser charge of aggravated assault, and in most jurisdictions there will be circumstances in which one or the other must therefore be chosen. *See* Michael H. Hoffheimer, *The Rise and Fall of Lesser Included Offenses*, 36 RUTGERS L.J. 351, 354 (2005).

into a ceiling. The ceiling has at least two sources. First, the criminal does not expect an infinitely long life, so a stretched out schedule like TH-1/AA-35/AB-95 is unlikely to produce greater marginal deterrence than does TH-1/AA-35/AB-50. Moreover, lawmakers will have prison conditions and prison construction costs in mind, so that widening the gap by pushing higher (and then pushing the penalty for maiming three witnesses yet higher) is a strategy with limits. The alternative is to widen the gap by lowering the penalty for the less severe crime. The analysis is not weakened by an assumption that the ceiling has been reached at AB-25, with thirty years reserved for murder perhaps, so that the way to further deter the armed robber from shooting is to lower the penalty for armed robbery alone. Perhaps TH-1/AA-10/AB-25 becomes TH-1/AA-5/AB-25, so that there is a significant deterrent to pulling the trigger, even if we are constrained by a ceiling and hold constant the probability of apprehension.[34] In some cases, this large gap is politically, or even morally, impossible. If the middle crime is especially heinous, as in the case of rape, marginal deterrence may simply be sacrificed in the interest of proportionality. Where armed robbery is concerned, however, the 1/5/25 schedule may be acceptable.

Note that the gap between theft and armed robbery narrows as the gap between armed robbery and aggravated battery increases,[35] and a thoughtful legislator might fear that thieves, as mundane as pickpockets, will now arm themselves and sometimes do great harm.[36] Given the ceiling, no gap can be widened without narrowing another. Fortunately, it is unlikely that offenders will easily substitute from pickpocketing or other thievery to armed robbery. They certainly will not do so on the spot. If marginal deterrence is about influencing an offender in the midst of a crime, the TH/AA gap is obviously less important than the AA/AB gap. Moreover, many pickpockets are experts in their trade and would not readily shift to armed robbery; other amateur thieves have no ready ac-

34. Most of the examples in the text assume a constant rate of apprehension. In reality, the gap between penalties usually understates marginal deterrence because the offender knows that the more serious crime will attract more police investigation. On the other hand, a more violent offender may scare away or even eliminate witnesses.

35. The punishment for theft depends on the value taken and whether T is a repeat offender. 720 ILL. COMP. STAT. 5/16-1(b). There are many different grades that increase the penalty. For example, if T steals property valued less than $10,000, it is a Class 3 felony punishable by two to five years imprisonment and a fine up to $25,000.

Robbery, however, is a Class 2 felony punishable by three to seven years. *Id.* § 5/18-1(c). The crime rises to Aggravated Robbery if T threatens with a weapon (even if he is lying about that fact), a Class 1 felony (four to fifteen years). *Id.*

The punishment for Armed Robbery depends on whether the criminal is armed with a dangerous weapon (Class X, six to thirty years), a gun (six to thirty years + fifteen years), discharges a firearm (six to thirty years + twenty years), or shoots the victim (six to thirty years + twinty-five years). *Id.* § 5/18-2(b).

36. The argument works best if these offenders underestimate the harm they will do with weapons.

cess to firearms. The most serious design problem is where a narrow gap might induce substitution toward more serious wrongdoing.[37]

The legislature's major problem is that this substitution effect contradicts the influence of marginal deterrence theory. If the schedule is adjusted to TH-1/AA-5/AB-25, there is the danger that armed robbery will become newly attractive to offenders. It may well cause a few criminals who are inclined to shoot to do less of that but then to commit many more armed robberies, as most of them plan to leave their non-submitting victims unharmed. More important, criminal activity may increase because armed robbery (sans shooting) is now underdeterred. Its ideal penalty of ten was decreased to five to widen the AA/AB gap and discourage shootings; but this move in support of marginal deterrence leaves an insufficient deterrent for the less serious of the two crimes, and there will be substitution in its favor. The first two dimensions of criminal deterrence are thus at odds with one another.

The important lesson here is that law needs to focus on substitution as well as escalation. It may, for example, be sensible to increase the penalties for both theft and armed robbery because criminals previously inclined to rob and shoot may now be likely to commit a few more armed robberies—in the course of which they are less inclined to shoot—but also to find mere theft, or even pickpocketing, more attractive.[38] In addition, other offenders may migrate to armed robbery because its penalty has been decreased. Taking these effects into account, the new schedule should perhaps be TH-2/AA-8/AB-25. Some of the marginal deterrence is sacrificed in order to make armed robbery ("AR"), and even theft, less attractive, but there is no reason to expect that the optimal schedule is a return to TH-1/AA-10/AB-25. In fact, the typical schedule of penalties dramatically widens the gap to discourage the use of a firearm, leaving fewer years to discourage the deployment of the firearm.[39] This is probably well-intentioned, but mistaken. There ought to be experimentation with a wider gap between the two serious crimes.

In some settings, there is probably no substitution of any kind.[40] Consider a plagiarist in a university community. If the penalty for a stu-

37. *See generally* Friedman & Sjostrom, *supra* note 3 (considering the two-crime cases as well as substitutions among more options).

38. Imagine that armed robberies, like most opportunities, are available to a given offender with a declining marginal productivity. The offender expects to gain 100 from the first, 95 from the second, and so forth, and has chosen to commit *x* armed robberies. The expected marginal product includes the risk of apprehension and the penalty for the crime. If law now increases the AA/AB gap, by raising the penalty for aggravated battery, then on the margin the offender will shoot less often and perhaps be prepared to commit more aggravated assaults without escalating to battery in the event that the victim does not submit. But the marginal productivity of armed robberies will fall, and mere thievery may also become an attractive option, as it is now the offender's best or second-best line of work, inasmuch as armed robbery—with some probability of aggravated battery—is now less profitable.

39. While using a gun greatly increases the penalty for armed robbery (six to thirty years + fifteen years), only five years is added for discharging the firearm and another five years for actually maiming or killing the victim. 720 ILL. COMP. STAT. 5/18-2(b).

40. Marginal deterrence is more straightforward where promises, rather than threats, are concerned, because substitution is unlikely. Thus, if a party to a contract hints at a coming breach, or breaches in a minor way in order to secure some benefit at the other party's expense, the law might or

dent's plagiarism were severe, it would be unlikely that plagiarists would think they had little to lose by stepping up to transcript forgery or the theft of rare books from the university library. The wrongdoing in question is fairly self-contained if only because the skill required for plagiarism does not help with the more serious crimes. Plagiarism may be like pickpocketing in this regard. Similarly, if the gap between sexual harassment and sexual assault on campus is increased by reducing the penalty for the former to improve marginal deterrence with respect to assault, it is improbable that plagiarists would switch to sexual harassment. If, however, there is a risk of increased sexual harassment because previously well-behaved students will now become harassers given the low penalty attached to that wrong, then it is risky to widen the gap between harassment and assault by lowering the penalty attached to the former.[41]

The offender who waves a gun while demanding a wallet can easily escalate to violence and aggravated battery; but because there is easy substitution back down towards mere threats and theft, reducing the penalties for, or even decriminalizing, threats may seem pointless. Reform in the direction of decriminalization may increase marginal deterrence for the more serious and violent crime, but people will gravitate toward threat-making if they believe they can often enough get what they want simply by threatening. But there is more to the problem than this tradeoff in two dimensions. In the next Section we introduce a third dimension and explain why a low sanction makes threats less credible because of the improved marginal deterrence; counterintuitively, these threats become less attractive to the offender. By lowering or even eliminating the sanction attached to threats, law can increase marginal deterrence *without* risking much substitution to the lower-level wrong of threat-making.

D. The Credibility Effect

1. Primary Credibility

If a rational or experienced criminal is often influenced by the gaps between penalties for crimes of different severities, then it is quite likely that he is also affected by his perception of his victim's regard for various threats.[42] If V finds a threat especially credible because V knows that T faces very little additional punishment by executing the threat T has visited upon V, then T will be more likely to make such a threat in the first place.[43] Correspondingly, when V finds a threat much less credible, be-

might not provide a remedy. If it does provide a remedy, then it deters the minor breach and threat, and it leaves less of a remedy to secure the larger project. But it is unlikely that a smaller penalty for the first breach changes the level of contractual activity.

41. And of course a better approach might be to widen the gap by increasing the penalty for assault, unless that is precluded by a ceiling formed by the university's capacity to punish without criminal law's protections and the involvement of the state.

42. *Cf. supra* Section II.B.

43. *Cf. supra* Sections II.A–B.

cause V knows that if T executes T faces a much greater punishment than if T does not, V will often refuse to submit.[44] In turn, because V will submit less often, T will be less likely to issue the threat in the first place—unless, of course, the penalty for the threat alone is so slight that T loses little by testing V. This is the heart of the primary credibility effect.[45]

Even where offenders can move fluidly from a threat to its execution, as they often can between aggravated assault and aggravated battery, the wider gap suggested by marginal deterrence theory has a powerful impact on the credibility of the threat—and this in turn *decreases* the amount of substitution to the threat. Note that the primary credibility effect is the third element to be considered in scheduling sanctions, but it is relevant only where one wrong amounts to a threat of completing a more serious wrong. The dimension of direct deterrence, reflected in the potential substitution of crimes with lesser penalties, never disappears because there will always be offenders who try to extract something from victims if they can threaten these victims at low cost. Even where the credibility of such threats is very low, some sanction is sensible, both because the threats impose anxiety and because some victims will submit. Still, the primary credibility effect should encourage law to attach low sanctions to threats. Doing so will create a larger gap and more marginal deterrence with respect to the more serious wrong of execution and, at the same time, will likely *reduce* threat-making. The idea is that there will be little substitution towards threats because the increased marginal deterrence reduces their credibility and profitability.

Both marginal deterrence and the primary credibility effect are a function of the gap between the expected punishment for the threat and the punishment for the more serious crime, but they are hardly the same thing. Marginal deterrence is about the offender's incentive to escalate from one wrong to another, as from AA to AB in our schedule. It is about T's calculation of the net cost of executing his threat, once he has issued the threat. In contrast, the primary credibility effect concerns T's incentive to issue the threat in the first place. It is derived from the probability of extracting a gain from V at either the first (threat) or second (violence) step, and thus from T's view of V's perception of the credibility of the threat. The primary credibility effect comes out of a dynamic analysis of threats. Marginal deterrence looks only to T's incentives and regards V as a passive player in the interaction. The substitution effect is somewhat dynamic in that it looks not only at T's responses to the schedule of punishments but also to other potential offenders' inclinations to engage in criminal activity. The primary credibility effect is yet more dynamic because it adds in the victim's perspective in the form of

44. *Cf. supra* Sections II.A–B.

45. It might be called the Incredibility Effect, because the idea is to use law and marginal deterrence to make offenders' threats less credible, but following convention we call it the credibility effect, as it is *about* credibility. The first (and only one, to the best of our knowledge) to point to what we call the credibility effect is Uri Weiss, *The Robber Wants to be Punished* (Hebrew Univ. of Jerusalem, Federman Ctr. for the Study of Rationality, Discussion Paper No. 685, 2015).

V's inclination to submit given the credibility of T's threat; moreover, it follows through by examining T's incentive to engage in threat-making in the first place, given V's likely response. For this reason, the primary credibility effect can be described as offering an *ex ante* perspective, while marginal deterrence operates not quite *ex post*, but *ex medio*—after the threat but before execution.[46]

The primary credibility effect can be understood as resolving the tension between marginal deterrence and the substitution effect. The former suggested a large AA/AB gap, for example, while the latter pushes in the other direction because offenders will substitute toward the lower-sanctioned wrong, AA. How do we know there is a net gain from increasing the gap? The primary credibility effect assures that the large gap is a good thing; by raising the cost of execution, it devalues the threat, or AA, itself. In the end, there will be little substitution towards something that is unlikely to bring about submission.

2. Secondary Credibility

As we explained in Part II, a threat is credible if the victim believes escalation to be likely in the event of nonsubmission. But it is secondarily credible the more V trusts T not to renew or execute the threat following V's submission. T wants V to believe that the threat is both primarily and secondarily credible, but low execution costs normally raise the probability of the first effect while lowering the likelihood of the second effect; low execution costs lead V to fear that submission will be followed by execution or another threat. If V submits, it is because V reasons that, once T obtains his property, T gains little from continuing violently.

Properly calibrated penalties can decrease both primary and secondary credibility and make threats even less valuable. We have already shown that law can decrease primary credibility by lowering the sanction for threats and increasing the gap between the sanctions for a threat and its execution. This may well be enough for most legislatures, but it is possible that law can do yet better by also decreasing secondary credibility. The counter-intuitive, and admittedly risky, strategy is to make the sanction for execution (or renewal of the threat) *following submission* low—perhaps even very low. If the sanction for post-submission execution (or

46. Imagine, almost as in Part II, that the sanction for a threat is eight, or alternatively one, and that the sanction attached to execution of the threat is twenty. Marginal deterrence is of course greater with the sanction of one for the threat. The gap between threat and execution is nineteen, or alternatively twelve. An offender who conveys a threat is presumably less likely to follow through and execute when the gap is greater. In contrast, the credibility effect focuses on the notion that the victim is less likely to submit when the sanction for the threat is one, and the offender is more likely deterred. This makes threat-making a less attractive activity to offenders, and there may actually be fewer threats because the sanction of one is perceived as costly given the fact that threats will rarely succeed. Meanwhile, if the sanction for the threat were eight, victims might be expected to submit with regularity because they know that unsatisfied threateners do not lose terribly much by proceeding with execution. If so, there will be more substitution to threats when there is a higher sanction attached to it. Counterintuitively, the 8/20 schedule might attract more criminals than the 1/20 schedule, and this is especially so if nineteen deters the greater wrong and twelve is inadequate for this task.

threat renewal) is low, secondary credibility will be low. V will know that submission is especially unwise because execution or renewal of the threat might well follow. In turn, T will see that V's submission is even less likely. The proposed legal strategy is unusual, to say the least. Ideally, threats will be reduced because the high AA/AB gap makes the threat incredible inasmuch as execution costs are high so long as there is a significant probability of T's apprehension. But if T threatens anyway, and V contemplates submission, the idea is to show V that submission might not satisfy T. If V is often convinced and does not submit, then in turn T will find threats even less profitable. In short, a wide AA/AB gap should convince V that T would not execute (as primary credibility is low). But when V nevertheless submits, it might be clever for the sanction for escalation to be low in order to weaken T's secondary credibility. The idea, again, is for T to see that V is unlikely to submit in the first place.

The potential for this unusual deterrence strategy is most apparent in a case like blackmail, where the victim has time to consider whether or not to submit to the wrongdoer's threat. V must fear that if she submits and pays T, T will simply repeat the threat. Imagine now that law penalizes blackmail but provides no penalty whatsoever for repeated blackmail. V will know that T loses nothing by repeating his threat (setting aside some increased chance of apprehension) and, in turn, V might now choose not to submit in the first round because the risk of repeated blackmail, with no end to required payments, convinces V to stand up to T in the first place. It is even possible that law could provide a substantial penalty for one-time blackmail, but provide that if a victim submits, then the victim is guilty of a crime or, more palatably, that the blackmailer is completely free of criminal liability once the victim submits to his threat. Again, the idea is to discourage submission by V so that blackmailers will find their craft unprofitable.

When a victim is in terrible fear and must act quickly, it is unlikely that a strategy aimed at reducing secondary credibility will work and, in any event, it is probably too fraught for legislative experimentation. Consider the rapist who threatens to maim or kill V if she does not submit. Law might focus virtually all of its deterrence at step one, so that aggravated criminal sexual assault is associated with a penalty of twenty-eight years, increased to twenty-nine in a case where a weapon is discharged or the victim is beaten, and then thirty (the presumed ceiling) for rape-murder. Current law more closely resembles a 10/20/30 schedule, which aims to deter escalation by the rapist—but at the expense of some low deterrence at the first step. In theory, an advantage of the 28/29/30 schedule is that it makes the threatener's implicit promise not to further injure the victim less plausible because the rapist does not face a significantly greater penalty in the event of escalation. But the implication is that law wants to encourage the victim to resist. And, even if it were somehow true that broad resistance would decrease the incidence of rape, the cost is too great to contemplate, and the likelihood that victims

can engage in cool calculations while being assaulted is low. Law might just as well tell the victims of a hijacked plane that they and their families will be unable to collect any damages unless they battle the hijacker to the death. It is possible that this sort of schedule of penalties for hijacking (15/25—but then zero in the event of submission) would discourage hijackers, but it would do so by sacrificing the lives of some victims. Such strategies seem more appropriate for trained military units than for civilians.

Returning to the earlier example of assaults aimed at gaining property, law can weaken T's secondary credibility by promising very little additional punishment to the criminal who uses violence following a victim's submission to an aggravated assault. If we imagine V to be a storekeeper, then it is plausible that V has thought about armed robbery in advance and, if V knows that law provides only a small gap between aggravated assault that ends in a property transfer ("AATH") and aggravated battery ("AB"), V may decide not to submit to an armed robber's threat. A narrow AATH/AB gap encourages V to disbelieve T's threat. In turn, an experienced T may be less likely to threaten in the first place. The legal strategy reflected in the narrow gap is risky because T might be tempted to escalate his crime at low cost to lower the risk of apprehension. If so, high marginal deterrence at the point of violence remains important. But it is plausible that the better goal is to reduce threats in the first place because these are the offender's high-profit crimes. If so, the strategy of diminishing T's secondary credibility by narrowing the AATH/AB gap is inviting.[47] On the other hand, if the most important thing is to deter violence, then the AATH/AB gap must be large.

E. Summary

Return now to our illustration where marginal deterrence considerations altered the initial schedule based on proportionality, and all that it might reflect, from TH-1/AA-10/AB-20 to TH-1/AA-5/AB-25. The substitution effect then modified this result to TH-2/AA-8/AB-25.[48] The primary credibility effect could surely nudge this back to TH-2/AA-6/AB-25, but a more interesting—if rather extreme—possibility is TH-2/AA-3/AB-25. Marginal deterrence matters where the second gap is concerned and, given the ceiling of twenty-five years for anything short of murder, deterrence is maximized if the penalty for AA, in this case

47. The legislature can decide separately whether the narrow gap should be at the low end, as with TH-2/AATH-4/AB-5, or at the high end, with TH-2/AATH-24/AB-25.

48. As conceded in the discussion *supra* Section II.B, it is possible that the marginal deterrence and substitution elements are perfectly offsetting, in which case AA-10/AB-25 is as good a guess as any for the optimal schedule. Recall that AB-20 increased to AB-25 to reflect marginal deterrence and the likely ceiling on penalties. TH-1 was modified to TH-2 in response to the substitution effect. The assumption of virtually no substitution between theft, or at least pickpocketing, and armed robbery means that the narrowing of the gap between TH and AA presents no danger. Finally, pickpocketing is not itself a threat of armed robbery or worse in the way that armed robbery is a threat of a shooting, so the credibility effect does not come into play; and TH-2 remains even as the AA/AB gap is adjusted to take account of the credibility effect.

armed robbery with no shooting, is about the same as that for theft without a weapon or threat.[49] The idea is to think beyond marginal deterrence to the danger of substitution to armed robbery. The driving force is primary credibility and the incorporation of the victim's perspective. The greater the AA/AB gap, the less credible is T's threat is because T is well deterred from executing on his threat—and the less attractive it will be for T, or any other offenders, to ply this threat. In turn, where it is most plausible that V and T have time to consider one another's incentives and responses, secondary credibility becomes more important, and a small sanction for execution after submission (or for rethreatening) is justified or, at least, deserves experimentation.

In reality, it is convenient to attach some extra penalty to a threat like AA, regardless of whether escalation follows. One reason is that if there is no marginal sanction at all, there will be too much opportunistic substitution, as offenders will simply have nothing to lose by threatening victims and hoping that some submit despite the incredible threats. Second, there is the political reality that victims will demand some punishment for their offenders. A third reason for having some sanction stems from the practicalities of law enforcement. A criminal who is apprehended in the course of an aggravated battery will insist that as long as his crime was not completed he should not be punished—and especially so if he did not even take the victim's property. There is too fine a line between armed robbery and "armed robbery with an attempted shooting" to rely upon.[50] Moreover, it would be costly and unacceptable to encourage police to delay apprehending repeat criminals in order to give the latter a chance to maim their victims so that they can be put away for twenty-five years.

49. A more complete analysis should take both submission and nonsubmission into account. When T threatens with AA, there are several possibilities: (1) V does not submit, in which case T either (1a) withdraws (a sequence we might call AAW) or (1b) escalates to AB. A large AA/AB gap is desirable to encourage (1a) rather than (1b); (2) V does submit, in which case T succeeds in gaining V's property (reflected in the sequence AATH), and then T either (2a) departs or (2b) proves secondarily unreliable and actually executes anyway (or repeats the threat). In this case, as discussed in the text, a large gap with respect to AATH/AB is good for discouraging (2b), but a small gap might be a remarkable way of defeating secondary credibility, and discouraging the threat, AA, in the first place. For example, if V submits, a schedule of AATH-4/AB-25 works to deter violence, while AATH-4/AB-5 discourages submission in the first place by sacrificing marginal deterrence. This is the critical choice for the legislature. The more it wants to influence V to join in crime fighting, the more a small gap might be used to reduce T's secondary credibility and thus encourage V's resistance. It is plausible that this choice depends on the legislature's choices regarding gun control. In a state where lawmakers believe they are fighting crime by allowing or encouraging storekeepers to arm themselves, the law presumably wants the storekeeper to resist in the TH/AA/AB scenario. It is even plausible that the structure of criminal sanctions should depend on whether the victim is armed.

50. Put differently, assault is normally thought of as attempted battery. At the same time, law often punishes attempts in order to have a remedy in place when a criminal is apprehended just before completing his intended crime. It is this function of the law of attempts that is unintentionally sacrificed when threats are completely decriminalized. One solution is to lower the penalty for threats, but not to go too far; in the text's example, TH-2/AA-6/AB-25 represents this position. Another is to classify threats, or aggravated assaults (with no battery), as the lowest class of felony or as a misdemeanors. AA-3 represents this solution.

The point of this Article is obviously not to draft a detailed criminal code but, for illustrative purposes, it is useful to settle on TH-2/AA-3/AB-25 as the schedule indicated by the combination of the proportionality principle, marginal deterrence, the substitution effect and, finally, the primary credibility effect. The penalty of three gives police a tool for apprehending a criminal just before a shooting, when there is a good chance that a threat has been issued. The low, but non-zero, sanction reflects the arguments and realities just described. And when secondary credibility seems relevant, law can experiment with a small sanction for escalation following submission.

IV. REVERSING THE CREDIBILITY EFFECT BY SINKING COSTS

A. The Sunk Cost Strategy

Marginal deterrence theory points to the danger of law's attaching a significant penalty to behavior that leads up to a serious crime.[51] In the case of a threat and the danger of its subsequent execution, the credibility effect is itself a function of marginal deterrence. It diminishes the value of threats to offenders. One way to think about the *ex ante* and *ex medio* perspectives is to describe the penalty that law attaches to a threat as a sunk cost. Once T brings on this sunk cost by threatening V, the sunk cost is irrelevant for marginal decision-making, and what matters is the expected penalty for escalation. The argument for lowering the sanction attached to threats is thus a way of undoing T's ability to sink costs and using the available deterrent to discourage execution and, in turn, make T's threat less credible in the first place.[52]

The sunk-cost perspective suggests that, even when all legal penalties are reserved for the final step, or law is simply out of the picture, T can enhance credibility by sinking execution costs.[53] Remarkably, this can

51. Shavell, *supra* note 2, at 345.

52. For an example that includes execution costs in addition to expected legal penalties, imagine that T threatens to harm V unless the latter transfers property worth nine. Suppose further that the expected cost to T, in the form of a legal penalty, is ten for the crime he threatens to execute, and four simply for threatening the crime. Apart from these legal penalties, T's costs are very low, perhaps one for cornering V and issuing the threat and then another two for executing it. T will gain the property worth nine either by coerced transfer or by taking it after he harms V. Criminal activity of this sort is highly profitable in the absence of law. T threatens V, spending one to gain nine, and V finds the threat credible. Even when "forced" to execute, T spends another two, and a total of three, to gain nine. But once law imposes penalties of four for a threat and ten for the (threat plus) completed crime, and often apprehends criminals, the criminal activity is unprofitable unless the victim often submits to the threat alone. At the point of threat, the marginal penalty for execution is ten minus four, or six, and when added to the execution cost of two, it pays for T to proceed in order to gain nine. If the threat alone is not penalized, and the entire ten is still attached to the completed crime, then T faces a cost of ten plus two, and crime does not pay. When it penalizes threats, law thus creates sunk costs and makes the completed crime—and thus the threat—*more* credible.

53. For an insightful analysis in one specific setting, see Lucian Arye Bebchuk, *A New Theory Concerning the Credibility and Success of Threats to Sue*, 25 J. LEGAL STUD. 1, 4 (1996) (suggesting that negative-value suits exist because the defendant knows that the plaintiff can proceed in stages, spending some resources in the first stage, and credibly threatening to spend more in the next, so that the suit can be credible even without actual expenditures). Note that Bebchuk assumes that the parties

be so even where it would seem that no credible threat is possible. The offender can sink costs, or stage the process of criminal wrongdoing, until it becomes credible that he will complete the crime. It is conceivable that V will foresee T's staging strategy and decide to submit even before T does much sinking of costs. This staging by T essentially reverses the credibility effect. Section III.D developed the idea that greater marginal deterrence for execution makes the earlier threat less credible, so that offenders are less attracted to threat-making because victims are less likely to be submissive. Here, the idea is that T can do the reverse; T engages in a kind of "homemade credibility effect," sinking costs in order to reduce the marginal cost of execution. This will enhance credibility (viewed from V's perspective), and make submission by V more likely. T will, in this way, increase the profitability of his threat.[54]

---

cannot affect the total costs of litigation or the expected value of the suit. The parties are also assumed to be capable of settling at any stage. We assume that settlement is not an option but instead that the threatener either gains all or gains nothing.

54. For an application of the idea advanced here about multi-staging to settlements, including a discussion of dividing surpluses, see William H. J. Hubbard, The Fourfold Credibility Problem in Law and War (2016) (unpublished manuscript). Although Bebchuk and Hubbard both picture the plaintiff as proceeding in stages, in much of civil litigation, sunk costs are not often in plaintiffs' control because discovery is usually more burdensome for defendants. Bebchuk, *supra* note 53, at 4; Hubbard, *see supra*.

Sunk costs and credibility are familiar concepts in warfare. Imagine a Country, T, that threatens to invade its neighbor, V, unless the latter cedes control of disputed territory. V finds the threat incredible because it knows that the cost of invasion to T is greater than the value of the territory at stake. Imagine that costs and benefits are similarly assessed by the parties, the territory in dispute is worth 100 to each, invasion costs 120 to T and would impose direct costs of 50 on V, and the invasion would be successful from T's perspective because it would control the territory in question and thus transfer 100 from V to itself. Finally, neither country gains anything when the other incurs costs. At the outset, T's invasion threat is not credible because it must spend 120 to gain 100. It tries to bluff because V's costs amount to 150 (the 50 from invasion plus the loss of the territory), but V can see that from T's perspective the invasion is too expensive. But imagine further that the 120 cost to T is the sum of 30 for mobilizing troops, another 30 for amassing equipment at the border once the troops are in place, and then 60 for the expected loss of life and equipment in an actual invasion. T might now proceed with mobilization. Following this first step, V will perceive that T need only spend 90 more to gain 100. Once T sinks costs, the threat becomes credible. If V rationally capitulates after T mobilizes, T will have spent 30 to gain 100.

Virtually every threat can be divided into stages, so this strategy for making a threat "appear" credible is of significance. As long as T is able to divide execution costs into stages that each costs less than the benefit to T from execution, V will capitulate. V can win only if (contrary to the illustration set out here) it can raise the cost to T of the final stage, as by defeating T's invading army. If not, V will comply in an earlier stage, perhaps even saving T the need to incur any costs at all, if each stage imposes costs on V and T's costs are of no benefit to V. Powerful bullies operate in such environments, the argument goes, and only rarely need to carry out their threats. The stronger party's threats are always credible, and something like concerted or principled reactions are required to thwart the skilled bully.

Finally, in some situations V can engage in symmetrical and offsetting staging. In the warfare example, V can influence credibility by taking steps that make capitulation impossible or at least less valuable to T. If V salts the earth, the disputed territory will be of lower value to T (as well as to V), and T's threat is made less credible. Working backward, T will not threaten because it perceives V's optimal strategy. Indeed, there should be no threats but only surprise invasions; the same is true for crimes, where V can destroy the value of what T seeks to acquire. In reality, and depending on technologies, V chooses a mix of precautions, including expenditures and sacrifices, to make it a less attractive target. The former can lead to an arms race, while the latter is analogous to a scorched-earth strategy. In turn, T can often increase the expected gains from execution in order to enhance credibility. T might, for example, build good highways up to the present border, thereby increasing the expected benefit of invasion. But, for many victims of crime, it will be impossible to anticipate threats and to undertake countermeasures.

B. Criminal Deterrence and Sunk Costs

In the context of criminal law, sunk costs derive from law's penalties more than they do from any direct expenditures by T. If the law contains a TH-2/AA-3/AB-25 schedule, then the threat made by an armed robber will not always be believed because V knows that the sanction T faces for execution is twenty-two. T, however, can reverse the credibility effect by sinking costs through strategic escalation. Thus, an armed robber might strike or pistol-whip V not only to frighten V into submission, but also to go beyond the lightly sanctioned category of aggravated assault, which is the threat staged in this case, to make his threat a battery. T would prefer not to shoot V because that risks raising the expected penalty yet further, inasmuch as some room was left for a murder charge. But T can make his threat increasingly credible by sinking costs in the form of expected penalties on the way to full execution. If V perceives this sunk-cost strategy, then V will find the initial threat more credible and will be more likely to submit at an earlier stage. In turn, the logic of the credibility effect now implies that T will find threat-making more attractive in the first place.

Where the credibility effect can be reversed so successfully, it may be a mistake to punish threats lightly. In these situations, marginal deterrence theory and the credibility effect suggest that law should choose its line of defense. In some cases, the best strategy is to apply nearly maximum marginal deterrence to the first act of violence committed by T, even if that is a threat of greater violence. The point is to defeat T's reversal of the credibility effect. Where the offender is able to sink costs, and where victims will submit because they perceive that staging is a winning strategy for the offender, the best deterrence strategy for law is to focus its penalties, and to do so at an early stage, where T sinks costs and where T's credibility is not yet enhanced. If armed robbers always threaten, then pistol-whip, and then shoot nonsubmitting victims—because at that point these offenders have little to lose—then law would do best by attacking pistol-whipping, which might be thought of as a point of execution rather than as an enhanced threat, though of course it is both.

Inasmuch as pistol-whipping is a battery, the TH-2/AA-3/AB-25 schedule may still be right.[55] The important point is not to create two modest gaps in an attempt to deter pistol-whipping and then also to deter a shooting; what is needed is one large gap to maximize marginal deterrence. If, for example, AB represents aggravated battery and ABF refers to the greater crime of aggravated battery with a firearm, as is true in many criminal codes,[56] then law should schedule one large gap either between aggravated assault and aggravated battery or between aggravated battery and aggravated battery with a firearm. The deterrent should not

55. The schedule might as well be TH-2/AA-25/AB-0, or equivalently, TH-2/AA-25/AB-25, where one who commits an aggravated battery is simply charged for the threat, or the lesser-included aggravated assault, except that there are cases where offenders simply shoot without first threatening.

56. *See* 720 ILL. COMP. STAT. 5/12-3(e) (2017).

be dissipated between the two gaps. TH-2/AA-3/AB-24/ABF is plausible, as is 25TH-2/AA-3/AB-4/ABF-25.[57] Unfortunately, current law is best described as dissipating rather than concentrating its deterrents.[58]

On the other hand, and perhaps unintentionally, judges have interpreted criminal statutes in a way that comports with the first of these two recommended schedules, and thus preserves some of the credibility effect. They have done this by interpreting "use of a firearm," or a similar statutory expression, to include not only the discharge of a firearm, but also pistol-whipping.[59] This practice goes against the traditional schedule in many states. Pistol-whipping, for example, is often distinguished from discharge of a firearm, leaving a large gap between "armed with a firearm" and the discharge of the firearm.[60]

Fortunately, there is another reason that strategic offenders cannot easily reverse the credibility effect—offenders who sink costs often give away their plans, allowing victims to defend and law enforcement to be more effective. For example, an offender who plans to tunnel into a bank vault must invest considerable resources in the hope of a large payoff. But staging, or sinking costs, is unlikely to work. Imagine a criminal who sends the following message to a bank: "Deliver $100,000 to me or I will tunnel into one of your bank branches and take at least twice that amount. The enclosed photos will show that my tunnel is well underway, and it is cheap for me to complete it and finish the job." On the one hand, the photos increase credibility because they show that the offender

57. The crime of kidnap presents a difficult case when it comes to parrying the criminal's sinking of costs. In a ransom kidnap, the offender often sinks costs before threatening physical harm to the hostage and of course law's heavy penalty for kidnap adds to the sunk cost. The threat of further violence is highly credible. One problem with lowering the penalty for the threat, or kidnap, is that the victim's family might pay the ransom without involving the police. In this case there will be substitution towards the underdeterred crime. Kidnap should be thought of as violence, and law's deterrent power should be focused on the kidnap even at the expense of some marginal deterrence with respect to inflicting physical harm on the hostage.

58. The dissipating nature can be observed by applying Illinois criminal law to this hypothetical. TH (theft, say pickpocketing a wallet) is punished with two to five years' incarceration. 720 ILL. COMP. STAT. 5/16-1(b). AA (assault) generates up to one year imprisonment. *Id.* § 5/12-1(b). AB (aggravated battery) is a Class 3 felony punishable by two to five years (though circumstances can increase this). *Id.* § 5/18-3.05(h). ABF (aggravated battery with a firearm) is a Class X felony punishable by six to thirty years. *Id.*

59. The pattern is likely to have less to do with insight about marginal deterrence and the credibility effect than with perceived toughness on criminals. *See* Montgomery v. State, 99 S.W.3d 257, 262 (Tex. App. 2003) (upholding defendant's aggravated assault conviction where the defendant exhibited the firearm while threatening to pistol-whip victim); Rose v. Commonwealth., 673 S.E.2d 489, 492 (Va. Ct. App. 2009) (holding that pistol-whipping during a robbery is use of a firearm in aggravated battery). The pattern extends to federal law. *See* Bailey v. United States, 516 U.S. 137, 148 (1995) (holding that "uses a firearm" means "active employment" and not mere possession of the firearm, under a drug trafficking statute, though active employment does not require the discharge of the weapon). After *Bailey*, 18 U.S.C. § 924(c) was amended to include mere possession. *See* Pub. L. 105-386, § 1(a)(1), 112 Stat. 3469 (1998) (codified as amended at 18 U.S.C. § 924(c)(1)(A) (2000)).

60. In Illinois, for example, use of a deadly weapon and "discharge of a firearm" are distinguished, and pistol-whipping would fall under the former given that it is not technically a discharge. The punishment for pistol-whipping, a Class 3 felony, is two to five years (but five to ten years with aggravating circumstances). 720 ILL. COMP. STAT. 5/12-3.05(h). Meanwhile, inasmuch as aggravated battery by discharging a firearm is a Class X felony punishable by six to thirty years, there is a large gap between pistol-whipping and shooting the victim. *Id.* § 5/12-3.05(g)(1)(C).

is not a mere extortionist without the capacity to tunnel into a bank. On the other hand, the bank can mount defenses at its branches, and police can begin to look for evidence of tunneling near the branches owned by the bank that is threatened.

Sinking costs is therefore a good strategy for a threatener only where the victim has no time to respond or where defensive maneuvers are very expensive. Criminal threats are most useful where they are specific and quick, as in the case of an armed robber, and it is in those cases that law must not help the criminal by enhancing the credibility of his threat with a heavy sanction. The discussion in this Part thus strengthens the argument in Section III.A. Where general threats are concerned,[61] one who threatens to harm a group in the future has allowed time for a defense to be mounted, so that the threat is not terribly credible. In any event, law normally attaches no sanction to a general threat so that maximum marginal deterrence also discourages the threatener's returning to harm the group.

V. CONCLUSION

This Article has introduced two new dimensions to deterrence in criminal law and suggested that these dimensions, or credibility effects, reinforce the importance of marginal rather than direct deterrence. We have pointed to the harm that law can do when it punishes threats severely. The criminal sanction operates as a sunk cost, reduces marginal deterrence, makes the escalation of criminal activity more likely, and makes crime itself more profitable. It is of course impossible to identify every situation where decriminalization or light sanctions for threats are unambiguously desirable, especially because much depends on the likelihood that victims will have time to reason about the likelihood that a threat will be executed. Law must be fairly general in its application, but criminals, victims, and their circumstances are heterogeneous. They will not uniformly assess the probability of apprehension, and that probability will, in fact, change with location, time of day, the criminal's skill, and the victim's behavior before, during, and after a threat or its execution. Criminals may also have disparate costs of escalation and different rates of success with their threats. Neither statutes nor after-the-fact judges can specify and identify these differences. It is necessary to generalize about the credibility of threats and the value of criminal penalties.

This Article has, however, demonstrated that lawmakers must resist the temptation to follow the proportionality principle all the way to the point where criminal law offers a long list of crimes of increasing severity accompanied by incrementally increasing penalties. This sort of schedule of criminal sanctions dissipates marginal deterrence and, most importantly, makes criminal threats more credible and more valuable to criminals. In many settings, the solution is to dramatically reduce the penalty asso-

61. *See supra* Part III.

ciated with a threat—even if the threat is as serious as one made by an armed robber who demands money and threatens to shoot if the victim fails to submit. Law should focus its deterrent power on the shooting rather than the threat, which is to say the aggravated battery rather than the aggravated assault. By doing so, the threat itself will become less credible and less valuable.

In some situations, criminals can sink costs, perhaps simply by beginning to execute their threats, in order to make their threats more credible, and thus more likely to gain submission from their victims. In these cases, law should focus its deterrent power on a single point, likely the first display of violence, even if that violence is a threat of a more serious harm. To do otherwise is to empower the criminal.

Where an ongoing criminal or terrorist organization is concerned, primary credibility is so great that deterrence is again best focused on the initial stage of wrongdoing. In practice, legislators devise special statutes to combat these threat-making machines. They might focus on "conspiracy," "racketeering," "terrorism," or even on membership in a recognized criminal enterprise or terrorist organization.[62] Common criminals are best deterred by lowering, rather than raising, the sanctions for criminal threats, but because organized crime is an expert at sinking costs and establishing credibility, the best deterrence might well be to aim at the organization's threats or first displays of violence, rather than its executions.

Most criminals use an element of surprise and cannot afford to sink costs and reverse the credibility effect to their own advantage because to do so is to give victims the opportunity to defend or call for help. In virtually all these cases, lawmakers must recognize that, when law's sanctions increase the credibility of offenders' threats, law itself is part of the problem. Penalties must be rescheduled to focus law's deterrent power on the danger of escalation or execution of threats. In this way, threats and other gateway crimes, though lightly sanctioned, will become less profitable to offenders.

62. *See* M. Cherif Bassiouni, *Effective National and International Action Against Organized Crime and Terrorist Criminal Activities*, 4 EMORY INT'L L. REV. 9 (1990); Zvi Joseph, *The Application of RICO to International Terrorism*, 58 FORDHAM L. REV. 1071 (1990) (assembling statutes and terms that are presently and potentially used to reach organized crime and terrorist activity).