The Six Elements of Legitimate Defense

George P. Fletcher
Jens David Ohlin

DOI: 10.1093/acprof:osobl/9780199757213.003.0004

Abstract and Keywords

This chapter pays close attention to the general principles governing the structure of self-defense in domestic legal systems and applied into the international arena. It defines the six essential elements of the doctrine of legitimate defense. These elements fall into two categories, three bearing on the nature of the attack and three on the requirements of permissible defense. In order for legitimate defense to be justifiable the attack must be: overt, unlawful, and imminent. The defense must be: necessary, proportional, and knowing or intentional in response to the attack.

Keywords: legitimate defense, international law, permissible defense

Our starting assumption is that the international law on the permissible use of defensive force should run parallel to domestic law. This point is anticipated in our discussion of the Kantian view that states are moral persons with rights and
duties, thus suggesting that parallels are properly drawn between individual and national self-defense.

We begin by paying closer attention to the general principles governing the structure of self-defense in domestic legal systems. In any particular legal system, whether we speak about self-defense, necessary defense, or legitimate defense, we are guided in our analysis by six structural elements. There is no statute or authoritative legal source expressing this consensus, but lawyers all over the world would readily acknowledge the same basic structural elements of a valid claim of self-defense or legitimate defense. We have touched on two of these elements already: necessity and proportionality. In addition, we have alluded to the other two: the requirements that the attack be imminent and that the defender respond knowingly or intentionally to the attack. Now we must assay these other issues systematically, with a view to the analogical relationship between domestic self-defense and self-defense in international law.

The six elements fall into two categories, three bearing on the nature of the attack and three on the requirements of permissible defense. First, the attack must be (1) overt, (2) unlawful, and (3) imminent. Second, the defense must be (4) necessary, (5) proportional, and (6) knowing or intentional in response to the attack. Depending on the particular issue, we shall reason sometimes from international to domestic law, sometimes from the domestic to the international.

1. The Attack must be Overt

Although the UN Charter overstates the issue, there is some wisdom in Article 51’s insistence that self-defense be invoked only when “an armed attack occurs.” The drafters meant to underscore the importance of an actual—and not merely perceived—breach of international peace and security. Something must happen in the real world, not just in the minds of those who fear an attack. The conventional, domestic law way to express this requirement is that there must be “overt acts” manifesting the intention to go to war.¹ Massing troops on the border, violating air space, closing shipping lanes, preparing missiles for attack—these are acts that are only preliminary to an actual attack but that manifest a hostile purpose.
The precise actions required are not so clear, either in theory or in practice. One standard of an actual, overt attack is prescribed by the prohibition in Article 2(4) of the unlawful use of force. Here the argument begins with a recognition of the value premises of the Charter, namely, “territorial integrity” and “political independence.” Apparently, any use of force that puts these values in jeopardy is unlawful. All states are required to abstain from these unlawful uses of force, with the exception of defensive action under Article 51.

Another approach is suggested by Article 39, which authorizes the Security Council to determine “any threat to the peace, breach of the peace, or act of aggression” and empowers the Council to take a range of measures to restore international peace and security. The relationship between this notion of breach and the defensive use of force under Article 51 is not made clear, but there is good reason to think that the exception for individually determined self-defense parallels the general role of the Security Council in deciding when actual or potential breaches of the peace occur.

It should be underscored, however, that all these standards are hotly contested. For political reasons, the General Assembly cannot agree on a definition of aggression. The main point at issue is whether the legitimate use of force is governed by the political criteria of resisting colonial, racist, and otherwise oppressive regimes. If this point is contested in discussions of aggression, the same problem attends the analysis of “unlawful use of force” under Article 2(4) and “breach of the peace” under Article 39. These issues turn out to be very subtle, but they raise serious questions about the foundations of legitimate force, which we take up in the following chapters. The important point at this stage of the argument is that any attack subject to a response of defensive force must actually, physically occur. There must be overt manifestations of aggressive intentions. The question of whether these manifestations need rise to the level of an “actual attack” will engage us later, when we consider the element of imminence.

The requirement of an overt attack in domestic law is not as obvious as it is in international law. The Model Penal Code has set the tone by emphasizing the belief that the actor is subject to an unlawful attack. There is no explicit requirement of an actual attack, provided the actor believes that one is about to
occur. There is an additional debate about whether the belief must be reasonable, but that debate runs orthogonal to the problem of where an attack must actually occur. The more fundamental divide lies between those who think that self-defense should be evaluated only against real and objective events in the world, and those who believe that these claims should be evaluated against an actor’s subjective beliefs about the world. After all, individuals can make grave mistakes in perception, and sometimes what looks like an attack is not one at all.

Here the Rome Statute is ambiguous. It provides that the actor must “act reasonably to defend himself or herself or another person ... against an imminent and unlawful use of force.” This seems to imply that there must at least be an overt threat to engage in the imminent and unlawful use of force. Take the case, recently reported in the press, of a U.S. Marine who enters a house in Iraq and finds a wounded insurgent lying on the floor. The Marine is afraid that the wounded man is a decoy lying in wait, and so he kills him. He claims that he has heard many stories of insurgents laying this sort of trap for American soldiers. Is he guilty of the war crime of murdering someone who should have been taken prisoner? Not if he acted in self-defense. He might have “reasonably” believed the wounded soldier was a decoy, but in our view that is not sufficient to invoke a claim of self-defense. The better reading of the statute is that the justification of self-defense depends on whether the defense is reasonably directed against an actual threat—not merely a perceived one.

Self-defense in jus in bello, therefore, follows self-defense in jus ad bellum. As you need an actual attack—at least the beginnings of actual aggression—to justify a defensive response under Article 51 of the Charter, so you need an actual attack, or the objective signs of one, to justify a legitimate response under the Rome Statute.

In this context, international law has something to teach the domestic understanding of self-defense. The writers on self-defense have become so engaged by issues of belief and intention on the part of the defender that they have ignored the requirement of a real-world event manifesting a hostile purpose. One reason we should require an overt act is that killing another human being is too grave an act to be justified
solely by the belief—even the understandable and reasonable belief—of the defender. There must be an event that the whole world can see to justify the use of force.  

2. The Attack must be Unlawful

It would be hard to find a national statute on self-defense that failed to require that the attack be unlawful. The Model Penal Code, the Rome Statute, and the German Penal Code are all clear on this point. The point of this requirement is to exclude self-defense against justified attacks (which are not unlawful), while permitting self-defense against unlawful attacks that might be excused on grounds of insanity, duress, mistake, and other considerations that deny the culpability of the aggressor. All legal systems are in accord about the importance of the distinction between justification and excuse. Justified actions are not unlawful; excused actions are. If justified actions are not unlawful, no one can invoke self-defense against them. This principle follows from the idea that those who are acting in self-defense are acting lawfully, which implies that they (metaphorically, of course) have the law on their side. In any particular dispute between incompatible interests, the law can be on only one side of the matter. The weight of the law comes down on the side of defenders and against aggressors.

Surprisingly, the element of unlawfulness is not expressly mentioned in Article 51 of the UN Charter, yet it is obviously implied. Consider the use of peacekeeping forces mandated by the Security Council under Article 42 to restore “international peace and security.” When such troops enter a country, can their intervention be considered an “armed attack,” triggering a right of legitimate force under Article 51? Technically, if you take the words at face value, perhaps so. But the point of the entire Charter is to deny this possibility. Troops dispatched by the Security Council (p.90) do not act unlawfully. The law is on the side of the Security Council. Their actions prevail over the inherent right guaranteed by Article 51.

3. The Attack must be Imminent

Most legal systems explicitly require an imminent attack. The common law was clear on this point, and German law refers to the requirement of a gegenwärtigen Angriff (present attack). The Rome Statute uses the word “imminent,” translated as well in the parallel official languages. Only the Model Penal Code deviates from this language by placing the entire burden
of analysis on the defense as being “immediately necessary ... on the present occasion.”

The requirement of imminence means that the attack has not yet occurred, but the defender cannot wait. This requirement temporally distinguishes self-defense from the illegal use of force in two ways. Retaliation against a successful aggressor is illegal because it is force used too late; a preemptive strike against a feared aggressor is illegal because it is force used too soon. Legitimate self-defense must be neither too soon nor too late.

It is fairly easy to see why the use of force may not be too late. If self-defense is exercised too late, after the threat has passed, the only way to justify it would be as punishment for the past attack, or, in the language of international relations, a reprisal or punitive response. But analyzing self-defense as punishment for a previous attack presents serious complications. The reasons are the same in both domestic and international law: because all citizens are of equal authority, no one is entitled to inflict on-the-spot punishment on another. Because all states are equal, no state is entitled to punish another. Indeed, Kant went so far as to claim that punitive wars were a conceptual impossibility. The very concept of punishment presupposes the authority to judge and condemn another, and because no state may do so relative to another, the effort to conduct a punitive action will always fail—not as a matter of fact, of course, but as a matter of principle. Even if states and individuals could logically punish one another, they should not. Just punishment presupposes a process of trial and deliberation; making a judgment on the spot and then executing it is hardly a legal process. Punishment is necessarily an institutional remedy, not one that individuals may take under their own initiative. So if force is used too late, after the threat has passed, it falls into the category of unjustified aggression.

(p.91) Explaining why the use of force may not be too early is more complicated. One would assume that when the use of force comes too soon, it should be described as preemptive, but we must exercise care in using this word. In standard cases of self-defense between individuals, lawyers rarely describe actual self-defense as preemptive action. When individuals face the threat of imminent attack—an aggressor
points a gun and the victim resists—the use of force is simply called justifiable self-defense.

Yet international lawyers routinely describe responses to imminent threats as preemptive self-defense. Michael Walzer follows this practice by describing Israel’s Six-Day War against Egypt in 1967 as, at best, preemptive action. He does not mean to condemn the Israeli action, however, and under the facts as they appeared to the world, it would have been difficult to do so. Egypt had closed the Straits of Tiran to Israeli shipping, massed its troops on Israel’s border, and secured command control over the armies of Jordan and Iraq. In the two weeks preceding the Israeli response on June 5, Nasser had repeatedly made bellicose threats, including threatening the total destruction of Israel. Most observers concur that Israel’s action was justified under the Charter of the United Nations. In a rare showing of solidarity in the United Nations, the Security Council—however quick it may be to chastise Israel’s policies—tacitly approved of its defensive action.
4. The Defense must be Necessary

As three elements define the kinds of attacks that may be resisted by defensive force, three factors circumscribe the permissible range of defensive action. Force is never justified unless it is necessary, but unfortunately, necessity is one of the most elusive concepts in our philosophical and legal vocabulary. Kant wrote, “Duty is the necessity of an action out of respect for the law.”\textsuperscript{16} He meant that duty-bound actions are dictated by duty in the same way that the natural world obeys the necessity of physical laws: if the sun shines on ice, then the ice necessarily, unavoidably, ineluctably melts.

When self-defense is described as necessary, the word “necessary” does not mean the same thing as our colloquial understanding of physical or literal necessity. Nonetheless, the criminal codes of the world refer to the necessity of defensive force, so a different sense of the word must be meant here. The German Penal Code is explicit; indeed, under German law, the defense is called “necessary defense,” or \textit{Notwehr}. The New York Penal Code requires that the actor “reasonably believes” the use of force “to be necessary to defend himself or a third person.”\textsuperscript{17} Interestingly, the Rome Statute says nothing about necessity in defending persons, but as to the defense of property the Statute demands that the use must be “essential,” either for human survival or for the success of a military mission. It seems safe to assume that “essential” and “necessary” function as synonyms in these statutes.

What, then, does necessity mean in this context? It is determined by the possibility of using a less costly means of defense: the attack must be repelled by the least costly method available, and there can be no reasonable alternative under the circumstances. In the \textit{Goetz} case, was there an effective response less drastic than firing the gun at the four feared assailants? Would it have been enough merely to show the gun in its holster? Or to draw and point the weapon without firing? Goetz had twice scared off muggers on the street merely by drawing the gun. This suggests that perhaps shooting to kill was not the least costly choice among the reasonable alternatives in the situation.

There are, of course, many arguments in Goetz’s favor.\textsuperscript{18} The uneven pace of the accelerating train made Goetz’s footing uncertain. During his initial exchange with Canty, he rose to
his feet and was standing in close quarters with his feared assailants. Showing the gun in the holster or drawing it would have risked the chance that one of the four young men would take the gun away and shoot him. Gauging necessity under the circumstances turns, in the end, on an elusive prediction of what would have happened if Goetz had tried this or that maneuver; short of shooting. In the end, the telling consideration is Justice Oliver Wendell Holmes’s wise aphorism, “Detached reflection cannot be demanded in the presence of an uplifted knife.”

A clearer case of unnecessary force is the March 1991 assault by twenty-odd Los Angeles police officers against Rodney King after they pulled him over and surrounded him in a parking lot in the San Fernando Valley. He had no apparent weapon. The police could easily have waited until he fell asleep, or they could have used a sleep-inducing gas to disable him. Instead, they went in with batons swinging. They claimed at trial that he had engaged in an aggressive motion known as the “Folsom roll,” so-called apparently because inmates at the Folsom Prison used it against their guards. The country was shocked by the videotaped brutality shown around the world. The early and excessive intervention seemed to be motivated more by a sense of wronged honor than the necessity of using force. Though the four indicted officers prevailed with their theory of self-defense in the state trial, the better interpretation of the officers’ motivation was that King had disrespected them by disobeying their orders. The officers were subsequently convicted in a federal civil trial.

Very few scholars pay careful attention to the requirement of necessity, even though it is a universal requirement of self-defense. In the past few years there has been considerable discussion about the duty to retreat when confronted with aggression on the street or at the workplace. The Model Penal Code introduced the duty to retreat, and in the wave of law reform that swept state legislatures in the 1960s and 1970s, many states added this duty to their penal codes. There is no duty to retreat at home, and some advocates of individual rights have argued that the same principle should apply on the street and at work. They are probably right, but changing the law on retreat does not substantially alter the requirement of necessity. Suppose that a handicapped aggressor, confined to a wheelchair, is approaching with a sword pointed right at you. Whether or not there is a duty to
retreat in this case, the use of deadly force would not be justifiable. Under the circumstances, it is not necessary to meet force with force. Stepping aside is sufficient to avoid the danger.\textsuperscript{\textit{24}}

In an international context, it would be difficult to argue self-defense if diplomatic means were available to solve the conflict. If Israeli delegates could have sat down with their Egyptian counterparts to negotiate a satisfactory end to the 1967 conflict, the resort to arms would not have been justifiable. In this context, Article 51 of the UN Charter includes an important qualification on the right of states to resort to defensive force; the “inherent” right of self-defense applies only “until the Security Council has taken measures necessary to maintain international peace and security.” A corresponding analogy in the domestic sphere might be a cowboy engaged in a shootout with bandits. When the sheriff’s men arrive on the scene, the cowboy would be expected to holster his gun and let the state’s officials take charge of the situation.

The literature of international law speaks of the requirement of necessity as well as proportionality, but no real sustained attention is paid to either of these criteria. There are good reasons for this indifference to these concepts. Necessity makes sense only relative to the prospect of avoiding and fending off an attack in progress, and the escape of the offender is typically included as an extension of the crime; preventing the offender’s escape is considered equivalent to preventing the commission of the offense, as in a famous 1920 German case in which an orchard owner shot and killed an escaping apple thief.\textsuperscript{\textit{25}} The shooting was ruled justifiable as necessary to prevent the successful commission of the crime—but of course there is an issue of proportionality here. For purposes of analysis, however, the two issues must be separated, and in fact the apple thief case is noteworthy precisely because it rejected proportionality as a limitation on the use of force.

It is important to note that in this context the criterion of necessity does not function in the same way as in international conflicts. In the nineteenth century, however, perhaps international lawyers thought it might be possible to prevent attacks from occurring. In Daniel Webster’s famous letter to Lord Alexander Ashburton about the sinking of the U.S. ship
Caroline, Webster claims that defensive force should be justified by “a necessity of self-defense, instant, overwhelming, leaving no choice of means, no moment for deliberation.” The British, he claimed, intervened prematurely against what they took to be American intervention on behalf of rebels in Quebec. Today, there is little chance of preventing armed attacks from occurring—except, of course, by striking while the threat is still imminent and not yet operative. If bombers are ordered to strike, if missiles are launched, there is little chance of warding off all these incoming agents of destruction. The most that a defending state can do is to hunker down, take the blow, and then strike in anticipation of the next attack. This is the typical response in international conflict: the defending state uses military force to destroy the military infrastructure of the aggressor nation and thus ward off successive attacks.

The rationale for legitimate defensive force was widely accepted for the U.S. invasion of Afghanistan after 9/11. An attack aimed at the Taliban made sense on defensive grounds, but surely we were not expecting another immediate attack as devastating as the destruction of the World Trade Center. However, this is not to say that there was no threat on the horizon at all. Rather, defensive force was designed to confront the possibility of future attacks that our enemies were plotting against us. There were, to be sure, a variety of future attacks that were in the planning stages, with organizers looking for funding, weapons, and trainees. But this should be carefully distinguished from a specific imminent attack on the horizon. The claim of defensive legitimacy was directed toward possible follow-up attacks in the future, but there were no signs of another attack in the coming days. The exact significance of this distinction between, on the one hand, current or imminent attacks and, on the other, future or planned attacks is discussed in chapter 7.

It is important to recognize that, in contrast to defensive action against aggressors in the domestic context, a military response to attack does not prevent the primary attack, but at best might forestall a follow-up use of force. The intent to avert secondary strikes blends readily with a motive of deterring subsequent aggression by the offending state. The first means of prevention is purely physical; the second seeks to change the motivation of the attacker. The United States and the United Kingdom have recently advanced the argument
that their use of force in Iraq was justified not as prevention in the physical sense, but as deterrence,\textsuperscript{26} thus starting down the slippery slope of preventive self-defense. The argument of physically neutralizing a not-so-likely second act of aggression is dubious, and the claim of a right to change motivation and deter is worse. If the argument of deterrence were valid, then military attacks could nearly always be justified; it would be too hard to tell the difference between illegal retaliatory attacks and lawful cases of deterrence.

Retaliation is in fact a frequent motive in conflicts, in both international and domestic law. Sometimes the impulse of retaliation is dressed up as a gesture of preserving honor. This was a widely shared sentiment after 9/11: a great power could not simply take a blow like that and not respond. But those who defend the use of violence rarely appeal directly to honor or admit that their purpose is retaliation for a past wrong. The more typical argument is that the individual or nation faces a recurrence of past violence, shifting the focus from past to future violence—from retaliation to defense against an imminent attack. In the case of Afghanistan, the motive of defending the honor of the United States by retaliating was probably uppermost in the minds of the nation’s leaders, yet no one dared admit it. The official rationale was, of course, self-defense against future terrorist attacks.

The criterion of necessity makes sense when we are talking about preventing a current attack or frustrating the commission of crimes, but not when the purpose is preventing or deterring successive attacks. What makes the use of force necessary in these circumstances? If bombing Afghanistan was necessary to prevent or deter future attacks—if that is the only factor that justified the bombing—then the argument surely applies against other threats that should be prevented or deterred as well. Under this theory, the analogous use of force against North Korea or Iran would also be justified, in addition to using force against any other nation that might attack us in the future, because we face the possibility of future attacks from them as well. Our actions would be “necessary.” Yet Afghanistan was viewed differently, largely because the elements of honor and retaliation made the argument of prevention easier to accept. But if the concept of prevention is doing the justificatory work in the argument, then it should not matter that we were already attacked once by Al Qaeda.

\textbf{(p.96)} The only thing that would matter is the possibility of
stopping future attacks, and those future threats come not just from Afghanistan and Al Qaeda but from a host of different regimes as well. At this point it would be difficult to determine who we can and cannot attack.

5. The Defense must be Proportionate
To understand the distinction between proportionality and necessity, let us return to the case of the orchard owner shooting the escaping apple thief. If we consider the prevention of the escape to be an aspect of preventing the crime, then of course it might be necessary to shoot. But shooting to kill, or even to wound, would not be proportional to the harm threatened. The distinction between necessity and proportionality means that in some cases the use of force might be necessary (no other way to stop the thief) but not proportional (the value of the apple is too small, relative to the taking of life). The disproportion between life and limb, on the one hand, and apples on the other is too obvious to require discussion. No legal regime that imposes a rule of proportionality would treat the shooting of the apple thief as justifiable.

Proportionality in self-defense requires a balancing of competing interests: the interests of the defender and those of the aggressor. Women may kill to prevent a rape. As the innocent party, the woman’s interests weigh more heavily than those of the aggressor. But although she may kill to ward off a threat to her sexual autonomy, the situation is more complicated when we ask whether she has a license to take life in order to avoid being kissed or touched. Of course, it matters what kind of touching we are talking about. If the only way she can avoid some truly innocent—but nonetheless unwanted—contact is to kill, that response seems clearly excessive relative to the interests at stake. Even if our thumb is on the scale in favor of the defender, there comes a point at which the aggressor’s basic humanity will outweigh the interests of an innocent victim, thumb and all. There is obviously no way to determine the breaking point, even theoretically. At a certain point our sensibilities are triggered, our compassion for the human being behind the mask of the evil aggressor is engaged, and we have to say, “Stop, that’s enough.”
The rule of proportionality makes sense, but not all legal regimes require it. The German provision does not explicitly require this balancing of the competing interests and has resisted the limitation of proportionality for decades. The traditional view was based on the Kantian theory that the (p. 97) defending party has the right to vindicate his autonomy, whatever the cost. Sometimes this is expressed as a categorical confrontation of right and wrong; in the German expression, *Das Recht braucht dem Unrecht nicht zu weichen* (The party in the right should never yield an inch to the party in the wrong). German writers in the Kantian tradition have reasoned that, as an “ethical matter,” the party in the right might stay the hand of defensive force, but there is no legal obligation to do so. In the post–World War II period, German courts have imposed this ethical duty as a legal requirement by importing a principle of private law called “abuse of rights” to help limit a right of defensive force that, according to the Penal Code, extended to all cases in which the use of force was necessary.

The European Convention on Human Rights explicitly limits the scope of self-defense to the defense of persons rather than property. Many American penal codes concur that the attack must represent a threat to the person as well as to interests in property. The New York Penal Law implicitly adopts these restrictions in the principle of proportionality built into its two levels of defensive response: the use of “physical force” and the use of “deadly physical force.” The former is permissible only to prevent the “imminent use of physical force” against oneself or against a third person; the more serious response, the use of deadly force, is permissible in specified cases where the threatened force is more serious. Of the scenarios enumerated in the New York Penal Law provision on self-defense, the ones that are most relevant to Goetz’s conduct are the threat to use deadly physical force and the attempt to commit a robbery.

The generally recognized right to use deadly force suffers from an ambiguous rationale. Breaking into a house need not, in and of itself, create a risk of harm to the homeowner or to anyone present on the premises. But in some cases such an attack on private premises might entail these risks, and therefore some legal regimes take the view that deadly force is permissible against burglary. Others reach the same conclusion by assuming that property in itself is worthy of the
same degree of defense as is applicable in cases of threats to human life. In the international arena, the parallel issue is the use of deadly military force to defend unoccupied territory. Suppose that there were no inhabitants in the Falklands when Argentina attacked. Would the British have been able to defend their interests—even at the cost of Argentine lives? This important issue of principle has received surprisingly little attention from international lawyers.

In applying the principle of proportionality in the international context, we have to distinguish rigorously between two types of cases. (p.98) In the one type, the one most like domestic self-defense, the use of defensive force actually prevents the attack. This is atypical in modern warfare, though one could imagine a line of troops blocking the advance of another line of enemy foot soldiers or cavalry. In these cases any necessary means would be permissible, just as in the Kantian model of self-defense. In the second type of case the attack has already struck its target, entailing casualties, and the issue is preventing further attacks or at least containing the continuation of the same attack. In this situation the problem of proportionality becomes critical. For instance, if a single Egyptian bomber reaches Tel Aviv and inflicts a hundred civilian deaths, Israel would undoubtedly respond, but nuking Cairo would be a disproportionate response. The issue of proportionality becomes acute in these cases because the normal requirement of necessity, as we have seen, does little work. Necessity makes sense in the first type of case, where there is some prospect of averting the attack altogether. Where the use of force is directed at presumed subsequent attacks, however, the issues of necessity and proportionality are both in play.

Though there are many different senses of the word “proportional,” international law pays almost no attention to this set of distinctions, but they are second nature to good domestic criminal lawyers. In fact, in the analysis of various claims of justification, the concept of proportionality recurs at almost every turn. There are four distinct sorts of proportionality. In cases of self-defense, the disproportion would take the form of clearly excessive force, such as shooting an apple thief running off with a single apple. Let us
represent this limited exception to self-defense by making it point 90 on a scale of 1 to 100 (Figure 4.1).

To understand Figure 4.1 and the following three illustrations, read the darkened portion as the protection that we afford to aggressors (e.g., apple thieves have a right to life). The blank portion of the graph represents the harm that can be done to an aggressor. Or, expressed from (p.99) the defender’s point of view, the blank portion represents the liberties that the law affords to defenders, for example, the right to respond with deadly force in all but the most extreme cases of clearly excessive force. It would be possible for Figure 4.1 to be all blank. That would be the case if proportionality were not recognized at all as a limitation on necessary self-defense, for example, if we were allowed to kill a thief for stealing an apple. The larger the darkened portion of the graph, the more significantly the principle of proportionality limits the use of force.

The concept of disproportionate force also applies to the traditional justification of necessity, but the meaning of “disproportionate” is entirely different. For example, in the language of the Model Penal Code’s provision on necessity (choice of evils), the person using force may act only “if the harm or evil sought to be avoided by such conduct is greater than that sought to be prevented by the law defining the offense charged.” Thus, if the disproportion in self-defense is represented as 90 on a scale of 100, the disproportion in cases of necessity would be signaled by the ratio of something like 45 to 55 (Figure 4.2). We get this ratio because the justification of necessity is allowed in cases where the evil sought to be avoided is greater than the harm inflicted by the use of force. However, notice that this standard does not require that the evil sought to be avoided be far greater. The evil sought to be avoided needs to outweigh the harm inflicted by the use of force only by a little bit before the defense is allowed. As one can see, this kind of disproportionality is far different from the disproportionality we considered above.
These two standards are often confused when assessing, say, the liability of commanders who send in bombers to attack military targets and cause disproportionate collateral damage to nearby civilians. Should the standard for quantifying disproportionality be the limited exception of self-defense, or the larger swath of illegality in necessity cases? This seemingly conceptual point makes a huge difference in the outcome of the

(p.100) case, because the ratios on the graphs are so different (i.e., 90 as opposed to 45). Do we assess this case by the self-defense standard of “clearly excessive force,” a permissive standard that benefits the defender, or by the standard of “greater harm” that we find in the law of necessity, a standard that allows less room for the defender to operate? Although it is clear that cases of collateral damage in war must be analyzed in terms of disproportionality, at first glance it is unclear if we should apply the version of disproportionality one finds in the law of self-defense or the one found in the law of necessity.

This is a subtle problem complicated by the doctrine of double effect, which holds that if the primary purpose of the action is legitimate (say, the pursuit of a military target), then collateral damage is acceptable provided that the damage is not disproportionate to the aim pursued. The underlying moral principle is that the action should be regarded as privileged because of the legitimacy of the dominant purpose. This helps answer the question of what kind of disproportionality is at stake in cases of collateral damage. It is not the sort that applies in cases of necessity as a justification, where the ends and the means are of equal moral value. Rather, we need to apply the notion of disproportionality that places the required harm near the 90 percent mark on our spectrum. It is for this reason that the Rome Statute requires “clearly excessive” harm in its war crimes provision on attacks that damage civilian targets.36
As if this situation was not sufficiently complicated, German private law also distinguishes between two different kinds of necessity in cases dealing with property interests. **Defensive necessity** arises when the defender must face a threat bearing down on him by a nonhuman actor.\(^{37}\) The classic example is shooting a rabid dog. The principles of self-defense apply to human actors who engage in aggression, but not to dogs, who are considered property and are not subject to the norms of the legal system. The interesting feature of defensive necessity (as applied in the case of shooting the attacking dog) is that, as in cases of self-defense, the defender is not required to pay compensation to the owner of the dog (or the victim, in cases of self-defense).\(^{38}\) Defensive necessity is also sometimes called “passive necessity,” to emphasize that the threatening property (in this case, the rabid dog) came to the defender, as opposed to a case where the defender actively seeks out the property that he ends up destroying.

Conversely, there are cases of **aggressive necessity** where the aggressor intrudes upon the property of another to save a higher interest.\(^{39}\) As typified by a famous American tort case, a ship owner moors his ship to a dock owned by another to avoid the devastating impact of an impending storm. The storm crashes the ship into the dock, causing minor damage. In this case, under both German and U.S. law, the ship owner is justified in mooring his ship, but he must still compensate the affected party for the loss of the dock. The reason for the difference is that he voluntarily sought out the dock to protect his ship.\(^{40}\) Consequently, the standard of necessity differs in this case from the standard applied in cases of defensive necessity. In both cases there is a limit imposed by the principle of proportionality, though each uses a different version of the principle. In the case of defensive necessity (e.g., shooting the attacking dog), the standard lies somewhere between necessity (as a justification in criminal law) and self-defense. It should therefore be represented as point 70 on the graph (Figure 4.3).
In the case of aggressive necessity, though, the standard of proportionality is far more demanding, precisely because the aggressor actively seeks out the property interest that is breached. This standard cuts off the permissibility of the action considerably earlier than the 50/50 line. Let us refer to this as point 30 on the graph (Figure 4.4). In other words, to actively intrude upon another person’s property and to cause damage, you have to have a very good reason. In cost/benefit terms, the standard for defensive necessity (shooting the attacking dog) states that the costs cannot be too great relative to the benefits. But the factor of disproportionality is higher than in cases of self-defense because the attacker is not a human being acting in violation of legal norms. In cases of aggressive necessity, where the actor intrudes upon the property of another and causes harm, the benefit must actually be much greater than the costs. It must be a matter of saving life relative to the cost of doing damage to property interests. (This makes sense because an intrusion against someone else’s property rights is always questionable, and therefore subject to tighter restraints.)

(p.102)
These differences reflect two value commitments. First, we should maintain a high respect for private property, even though we have less respect for property if it is a source of risk to others (i.e., the rabid dog). However, if the aggressor selects the victim and deliberately invades the property of another (as in choosing to dock without permission in order to protect the ship), the invasion against the rights of another is more palpable. Second, passive victims of threats have greater rights than those whose body becomes the instrument of harm. If a fetus growing inside a mother’s womb threatens her life, the principle of defensive necessity would apply to justify abortion of the fetus—provided, of course, that the death of the fetus is regarded as being within the realm of proportionality. In this scenario, it is of critical relevance that the fetus’s moving in the mother’s domain is described as a threat. If the fetus were in its own domain, apart from the mother, it would be much more difficult to justify killing the fetus to generate a benefit, even a life-saving one, for another person.41 There is another way of expressing these distinctions. If a thing is “tainted” because it generates a risk to others, we generally grant it less value in the balancing process, and this lesser value is reflected by the fact that it becomes possible to apply defensive necessity (as opposed to necessity as the lesser of two evils or aggressive necessity).42 That being the case, we are left with a serious problem: Why, as a matter of principle, should it matter whether or not an object is a source of risk in a modern rational system of criminal law? Literature in this area has not yet generated a convincing answer.43 Still, though the distinction between aggressive and passive necessity is not recognized in either Anglo-American criminal law or international criminal law, it remains a valid option that appeals to some scholars.44

(p.103) 6. The Defense must be an Intentional or Knowing Response
The preceding two characteristics of the defense—necessity and proportionality—speak to the objective characteristics of using force under the circumstances. To assay these issues, we need not ask any questions about what the defending individual or state knew about the attack and the defense. In the typical case, the defender is fully aware of the circumstances of the attack and chooses to respond with defensive force. But, to take a variation on the Goetz case, assume a white passenger is surrounded by four black youths on the subway. No one has made a threatening move, but in fact the youths have surreptitiously drawn their knives and are poised to strike against the “other” in their midst. All of this is
unknown to the passenger, but at that moment—for his own reasons—he pulls his gun and fires with the sole aim of inflicting harm on the four black youths surrounding him. Only after they have been disabled is it established that they had knives in their hands and were presumably ready to attack. In this variation of the case, could Goetz invoke self-defense on the ground that his act did, in fact, frustrate the attack? It would be a de facto act of self-defense, even though Goetz had his own reasons for shooting.

The consensus among Western legal systems is that to legitimately invoke self-defense, the defender must know about the attack and act with the intention of repelling it. Why should Goetz receive the benefit of a justification if he acted maliciously, without knowledge (or fear) of an attack?

Surprisingly, some leading scholars think that in a case of criminal homicide, the accused should be able to invoke self-defense even if he does not know about the attack. The scenario would look like this: criminal A plans a murder and starts to commit the crime. Unbeknown to him, though, his victim B was about to attack him first with a weapon. Had criminal A known about this attack, he surely would have been justified in using defensive force to stop it. The only problem is, criminal A knew nothing about the attack and instead killed his victim in cold blood. Knowledge of the coming attack played no role in his reasoning at all. However, some legal scholars argue that such a criminal should be able to plead self-defense anyway. They argue that, if you cannot be guilty of homicide for killing someone who is already dead (no matter what your intent), then you should not be guilty of homicide for killing an aggressor (no matter what your intent). No harm, no crime—and there is arguably no harm in killing an aggressor. To put the best possible face on this position, we might think of the injured aggressor’s right to complain. Suppose the police arrest someone without probable cause but, though the police acted illegally, the suspect is in fact guilty. Should he be able to complain about the police arrest when they fortuitously got it right? In the context of the Iraq invasion, suppose that the United States gets lucky and finds weapons hidden deep in a cave in Iraq. If we assume that the United States did not have a reasonable expectation of
finding weapons in the first place, should its subsequent good luck justify its conduct?\textsuperscript{46}

There is a tendency to argue that the guilty party (e.g., Iraq, if it had WMD) has no right to complain—in Iraq’s case, to the international community. This might be true, but the community as a whole may also condemn any use of force that is not motivated by the proper reasons. It seems counterintuitive to call someone an aggressor “in mind,” and then to allow him to benefit from discoveries made after the intended aggression is complete.

Scholars of criminal law who favor giving the aggressor the benefit of this good luck argue that he should nonetheless be guilty of an attempted killing. In such a case, the killer was acting for malicious reasons, though it turns out, after subsequent investigation, that he would have been justified by self-defense. On this view, the killer is neither fully culpable, like a murderer, because of the existence of objective circumstances that would have justified self-defense. But on the other hand, the killer is not fully absolved, like a bona fide defender, because the objective circumstances were completely unknown to him and he acted for malicious reasons. The killer’s culpability might lie somewhere in between, which is why some scholars are tempted to shove this case into an intermediate category like attempted homicide. He attempted to commit homicide, but in some sense he was frustrated from meeting the legal requirements of a murder. There is some support for this view in German law, and Paul Robinson has defended it with some success in the American literature.\textsuperscript{47}

This theory could be applied in the international sphere. That is, in the hypothetical case where the United States fortuitously discovers WMD after attacking Iraq, it would then be guilty only of “attempted aggression” against Iraq. Although the phrase sounds incongruous, the theory should be evaluated. One might wonder how the United States could be guilty of attempted invasion if, in fact, it had already landed over a hundred thousand troops on Iraqi soil and caused widespread death and destruction. Attempts are cases of trying and failing to realize the criminal plan. How can it be said that, in our variation of the Goetz case, the shooter tried but failed to hit the youths, when they are in fact lying bleeding on the \textsuperscript{(p.105)} ground? If weapons are found in
Iraq, can we really say that the United States (which did not have an *ex ante* justification for the invasion) is guilty only of an attempted invasion of 250,000 troops in Iraq?

The argument for Robinson’s position, amended for international law, is very subtle. In his scheme, the crime would not be invasion per se, but *unlawful* invasion. The aggression is supposedly not unlawful if weapons of mass destruction are found. The United States might have the intent to commit an unlawful invasion—that is, to invade regardless of whether weapons are found—but that becomes impossible under the actual circumstances of the case. If the weapons are there, it supposedly would become impossible to commit an unlawful invasion. Therefore, the United States is guilty of attempting to commit an unlawful act. This thesis testifies to the ingenuity of the legal mind, but it runs afoul of our understanding of terms like “trying” and “attempt.”

A better rule might be that self-defense is a privilege that can be properly exercised only by people who know the relevant facts. The Model Penal Code stresses the element of belief as a condition for any claim of justification. The New York Penal Code explicitly requires that the actor “reasonably believes” that the circumstances warrant defensive action. The implication is that the defendant cannot take advantage of circumstances beyond his knowledge and intention. The ramifications for the *Goetz* case are far-reaching. Ramseur and Cabey were carrying screwdrivers in their coat pockets; some media even reported that the screwdrivers were sharpened. The reports confirmed what many people wanted to believe: that these four kids were about to mug Goetz and therefore his response was justified as self-defense. Yet there was no evidence that Goetz knew about the screwdrivers. Not one of the four victims pulled a screwdriver from his pocket, either before or during the shooting. And if Goetz was not aware of the screwdrivers, then they had no bearing on his claim of justification. They had no more legal relevance than would a secret plan to kill Goetz because he was white.

The same rule should apply in the international context. The United States cannot claim justification for invading Iraq unless it is motivated by facts “reasonably believed.” The claims about WMD must have been believed, and as we shall see later, these beliefs must have been based on public evidence accessible to all. The relevant question, therefore, is
not justification ex post, or after the action, but justification ex ante, on the basis of facts held to be true before the invasion. The real question in Iraq is whether the Bush administration had enough credible, publicly revealed evidence to believe they would find something that had eluded (p.106) the United Nations inspectors. President Bush might have believed that he would find something, but, as in the Goetz case, the real problem was whether his belief was reasonable on the basis of the evidence publicly available. (Precisely why the belief must be reasonable and based on public evidence are matters that we take up in greater detail later.)

The requirements of imminence, necessity, and intentionality are found in numerous legal systems, as well as in the international law of armed conflict; proportionality, however, is a more contested requirement and deserves greater development. In any event, what we have offered in this chapter is not the final word on these concepts. Rather, we have simply introduced their broad outlines and explained their importance for any good theory of self-defense. The full richness of these concepts will be brought forward in full color in the following chapters, as we apply our theory of legitimate defense to the most pressing problems of international law: aggression, humanitarian intervention, and preemptive war.

Notes:


(3.) MPC § 3.04.

(4.) This is the debate in the Goetz case, which was resolved in favor of the “objective” standard of reasonableness in People v. Goetz, 502 N.Y.S.2d 577 (N.Y. 1986).

(6.) Rome Statute § 31(1)(c).

(7.) See Menendez George P. Fletcher, With Justice for Some: Victims’ Rights in Criminal Trials (Reading, Mass.: Addison Wesley, 1995), at 146-47.

(8.) This requirement of public evidence is discussed in greater detail in chapter 7.

(9.) MPC § 3.04; Rome Statute § 31(1)(c); StGB § 32.

(10.) For a defense of this analysis, see George P. Fletcher, The Right and the Reasonable, 98 Harv. L. Rev. 949 (1985).

(11.) The phrase “imminent and unlawful use of force” in 31(1)(c) appears in the French translation of the Rome Statute as recours imminent et illicite à la force, and in the Spanish as un uso inminente e ilícito de la fuerza.

(12.) MPC § 3.04(1).

(13.) Kant, Legal Theory at 347 [Gregor at 153].

(14.) See Brownlie, International Law and Use of Force by States 257–61, 366–67 (mixing up the terms “anticipatory action” and “preventive action” with the claim that the force must be directed toward an “imminent attack”). See the discussion of the Caroline incident below in the text.

(15.) Walzer, Just and Unjust Wars, at 81.

(16.) Kant, Moral Theory, at 4:400 [Gregor at 13].

(17.) See N.Y. Penal Law § 35.15(1).

(18.) For further discussion, see Fletcher, A Crime of Self-Defense at 29.

(19.) Brown v. United States, 256 U.S. 335, 343 (1921) (Holmes, J.).

(20.) For extensive analysis of both trials, see Fletcher, With Justice for Some at 37-68.

(22.) MPC § 2.04(3)(b)(ii) (“actor knows that he can avoid the necessity of using such force with complete safety by retreating”).

(23.) MPC § 2.94(3)(b)(ii)(1) (no duty to retreat from home or place of work).

(24.) See Mitchell Berman, Justification and Excuse, Law and Morality, 53 Duke L.J. 1, 13 (2003). In this otherwise thoughtful article, Berman ignores the requirement of necessity in self-defense cases and therefore claims, incorrectly, that the defense applies without moral justification, as, for example, in a case in which the defender could easily avoid the attack by stepping aside but does not.

(25.) Decision of September 20, 1920, 55 RGSt. 82 [hereafter cited as Decision of September 20, 1920].


(28.) This ethical duty is expressed in the principle Rechtsmissbrauch or “abuse of law.” See Kommentar § 32, point 46, at 426.

(29.) For a detailed discussion of abuse of rights, see chapter 5, section 4.

(30.) ECHR, Art. 2(2)(a).

(31.) NYPL § 35.15.

(32.) This view is taken in Talmudic law, which reasons that the homeowner will resist the burglar, which in turn will escalate the conflict and lead to risk of deadly attack. Babylonian Talmud, Tractate Sanhedrin 72.

(33.) See the State of Colorado’s “Make My Day” statute, Colorado Revised Statutes § 18-1-704.5 (2004).
(34.) For some doubts on the legitimacy of defending unoccupied territory, see Rodin, *War and Self-Defense*, at 134–37.

(35.) *See, e.g.*, Brownlie, *International Law and Use of Force by States*, at 261–64 (discussing proportionality as though it were synonymous with necessity).

(36.) Rome Statute § 8(2)(b)(4) (making it a war crime to “intentionally [launch] an attack in the knowledge that such attack will cause incidental loss of life or injury to civilians or damage to civilian objects or widespread, long-term and severe damage to the natural environment which would be clearly excessive in relation to the concrete and direct overall military advantage anticipated”).

(37.) German Civil Code (BGB) § 228.

(38.) *Id.* There is an exception for cases in which the defender brings about the attack by, say, letting the dog out of confinement.

(39.) German Civil Code (BGB) § 904.


(41.) For a difficult and wrenching case, see the problem of conjoined twins *In Re: A (children)*, 4 All E.R. 961 (2000). The court decided in favor of separating the twins to save the one who had the substantially better chance of survival.

(42.) The same principle is reflected in the law of eminent domain. Objects that are a source of danger to others may be destroyed without paying compensation. There is an apparent analogy to the hypothetical of the rabid dog in passive necessity. *See United States v. Caltex* (Philippines), 344 U.S. 149 (1952).


(44.) Notably Byrd, *id.*


(48.) The more general problem is the theory of legal impossibility, which is addressed systematically in Fletcher, *Rethinking Criminal Law*, at 174–84. The essence of the critique is that the factor “unlawful invasion” is not relevant to the U.S. motivations, and therefore you cannot rationally attribute to the United States an effort to bring about an unlawful invasion.