Chapter 7
The Ethics of Cyberattack

Steven P. Lee

Abstract The internet has made it possible to do damage at a distance by the use of networked computers. A deliberate act doing such damage may be referred to as a *cyberattack*. My concern in this essay is the ethics or morality of cyberattack as a part of war. The morality of war or military attacks in general is judged in terms of just war theory, which examines war in its two aspects, the morality of going to war (*jus ad bellum*) and the morality of conduct in war (*jus in bello*). I examine the morality of cyberattacks in each of these areas. My conclusion is that, while the use of cyberattacks is a novel form of conflict in many ways, its ethical dimensions can for the most part be understood in terms of the traditional categories of just war theory. There remains, however, an important aspect of cyberattack that may carry us beyond the limits of traditional just war thinking about war.

The internet has made it possible to do damage at a distance by the use of networked computers. A deliberate act doing such damage may be referred to as a *cyberattack*. In the words of one study: "Cyberattack refers to deliberate actions to alter, disrupt, deceive, degrade, or destroy computer systems or networks or the information and/or programs resident in or transiting these systems or networks" (National Research Council 2009, p. 1). Cyberattacks (also called computer network attacks) are a species of information operations. Cyberattacks may be carried out for a variety of purposes, just as ordinary (non-cyber) attacks may be. For example, there is cybercrime consisting of cyberattacks, as there is ordinary crime consisting of ordinary attacks of various sorts. When a series of ordinary attacks is carried out by a state against the interests of another state, this is sometimes a conventional, shooting war. Cyberattacks too may also be carried out by states against the interests of other states. Randall Dipert notes that cyberattacks may be "coordinated by the central commands of governments (or other political organizations), and [may be] directed at another country's governmental and military information systems, or at its commercial or infrastructure information systems for political purposes" (Dipert 2010,

S. P. Lee (✉)
Hobart and William Smith Colleges, Geneva, USA
e-mail: lee@hws.edu

L. Floridi, M. Taddeo (eds.), *The Ethics of Information Warfare,*
Law, Governance and Technology Series 14, DOI 10.1007/978-3-319-04135-3_7,

p. 385). When a state is, in this way, directing cyberattacks against another state, the result may be a *cyberwar*.[1]

My concern in this essay is the ethics or morality of cyberattack. The morality of war or military attacks in general is judged in terms of just war theory, which is the millennia old intellectual tradition in the West for assessing war in moral terms. Just war theory examines war in two respects, the morality of going to war, which is traditionally referred to as *jus ad bellum*, and the morality of fighting in war, referred to as *jus in bello*. So an ethical examination of cyberattacks should consist in considering cyberattack from each of these perspectives. In the first section, I examine some general issues on the nature of cyberattacks and cyberwar. The second section is devoted to the consideration of cyberattacks from the *ad bellum* and *in bello* perspectives, and the third section raises some further ethical issues raised by cyberconflict. My conclusion will be that while cyberattack is a novel form of conflict in many ways, its ethical dimensions can for the most part be understood in terms of the traditional categories of just war theory. There remains, however, an important aspect of cyberattack that may carry us beyond the limits of traditional just war thinking about war.

7.1 Section I

There is no doubt that cyberattacks can have an *operational role* in conventional war, that is, they can be (and have been) used in conventional warfare, for example, to disrupt the opponent's military communications (National Research Council 2009, p. 2). But some have argued that because cyberattacks do little or no damage in the physical world (as opposed to cyberspace), they are not sufficiently destructive by themselves to initiate or constitute a war. We may express this point by saying that stand-alone cyberattacks are not acts of war and that there can be no such thing as a cyberwar, understood as a war consisting largely or exclusively of cyberattacks.[2] Whether there could be a cyberwar in this sense may seem a merely verbal matter, but the answer to the question has important normative implications, which makes it worth our consideration.

[1] I do not have much to say in this paper about the use of cyberattacks by non-state agents, because, as I claim later, the likelihood that such attacks could rise to the level of acts of war is not significant.

[2] As a point of comparison, note that some would, for a very different reason, deny that there could be a nuclear war, understood as a major war consisting largely or exclusively of nuclear attacks. They would argue that "nuclear war" is a misnomer on the grounds that it must be possible for a war to have winner in the traditional sense, which a large-scale nuclear conflict would not have. There could not be a "nuclear war" because nuclear attacks are too destructive, while there could not be a "cyberwar" because cyberattacks are insufficiently destructive.

On a standard definition, war is the use of armed force for political purposes by one state in a large-scale conflict with another state.[3] International law makes the use of armed force a necessary condition for war (Schmitt 1998–1999). The main objection to the idea that there could be cyberwar is that cyberattacks do not conform to this definition. Cyberattacks, it is claimed, do not involve a use of armed force. No force is used in a cyberattack, and computers are not "arms" (Dipert 2010, p. 396). Ordinary war takes place in the physical world involving *kinetics* and physical damage. A cyberattack by itself kills no one; it is a matter of disruption rather than destruction. Note that in the definition of cyberattack in the opening paragraph, the harm that cyberattacks do is to cyber networks themselves and the data they contain, not to anything in the physical world apart from computer hardware. In addition, cyberconflicts take place in "cyberspace," which is different from physical space. In this sense, a cyberattack involves no crossing of borders, which are markers in physical space, and no violation of sovereignty understood as territorial integrity.[4]

One proponent of the argument that cyberconflict is not war is Thomas Rid. He argues that a cyberconflict is not a war because it fails to satisfy the three conditions necessary for war, conditions similar to those in the definition above. Rid argues that a war must be lethal, instrumental, and have a political goal. He argues that the stand-alone episodes of apparently state-sponsored cyberattacks to date have all been examples of subversion, espionage, or sabotage. No act in these categories, he claims, satisfies the three conditions, so none of these episodes has been by itself an act of war (Rid 2011, p. 2). The term "cyberwar," he asserts, involves a metaphorical usage of "war," as in the phrases "war on obesity" or "war on cancer." He suggests that there is a spectrum of activities between crime at the one end and conventional war at the other. State-sponsored cyberattacks with a political motive reside in the middle of this spectrum (Rid 2011, p. 3). Like other examples of subversion, espionage, or sabotage, they are the sorts of acts states may commit against each other outside the context of war. Others have, like Rid, made the claim that cyberattacks, taking place in cyberspace, are nonlethal (Bayles 2001, p. 47). Without the kinetics of regular war, there is little or no physical damage.

Were stand-alone cyberattacks not acts of war, a normative implication would be that they would not be covered under the law of war. While they might still be covered under other aspects of international law, these aspects might be weaker or more controversial in their application. The result might be that states would be substantially free to pursue a broad array of cyberattacks without contravention of their obligations under international law (Schmitt 1998–1999, p. 935; Schmitt 2002, p. 396).

[3] For more general purposes, revisions would have to be made in such a definition to account for civil war in its various forms. Later I will address the role of non-state agents in cyberconflict.

[4] When a series of cyberattacks were aimed at Estonia in 2008, NATO refused Estonia's request to invoke the collective self-defense provision of the NATO treaty on the ground that its sovereignty had been violated, stating that "a cyber attack is not a military action" (Lucas MS, p. 9).

But the argument that stand-alone cyberattacks are nonlethal and largely harmless, and so cannot be acts of war, is not sound. The argument depends on one or another of two implausible premises (Rid seems to rely on both of them). The first questionable premise is that we should expect that cyberattacks will do little physical harm because they have to date done little physical harm. The second relies on a cramped understanding of what counts as an effect of a cyberattack. Regarding the first premise, while it is true that cyberattacks have to date not done much physical damage, there is no reasonable expectation that the will continue in the future. The military application of cyber technology has not yet matured. The recent public concern about cyberattacks is due precisely to the reasonable belief that in the future cyberattacks will be able to do a great deal of harm. Indeed, this has already occurred. The Stuxnet computer worm has reportedly done serious physical damage to centrifuges being used by Iran to enrich uranium. Referring to this cyberattack, Michael Hayden, former head of the American CIA said, "Previous cyberattacks had effects limited to other computers…. This is the first attack of a major nature in which a cyberattack was used to effect physical destruction." He concluded: "Somebody crossed the Rubicon." (Quoted in Sanger 2012) The second premise relies on a bogus distinction between direct and indirect effects. The claim that cyberattacks are inherently nonlethal is like a claim that shooting a rifle is nonlethal because all it does is send a projectile through the air. Cyberattacks have the potential to do a great deal of damage (albeit indirect) in the real world, including the loss of human life. As the US government notes: "Critical life-sustaining infrastructures that deliver electricity and water, control air traffic, and support our financial system all depend on networked information systems" (Whitehouse 2011, p. 3). When such systems are deliberately attacked, the damage can be severe. While not all cyberattacks would have lethal effects, many would have lethal effects, and, more importantly, many would be intended to have lethal effects. Joseph Nye notes: "Major states with elaborate technical and human resources could, in principle, create massive disruption as well as physical destruction through cyber attacks on military as well as civilian targets" (Nye 2011, p. 21).

To give an example of one possible future scenario for a series of cyberattacks on the United States, consider the case sketched by authors William Clarke and Robert Knake (Clarke and Knake 2010, pp. 64–68). Fires have erupted at oil refineries across the nation, major gas pipelines have exploded, and toxic clouds of chlorine gas have been released from chemical plants. Air traffic control systems have collapsed, leading to multiple airline crashes, and train routing systems have failed, leading to multiple crashes and derailments. Signal lights have failed, resulting in accidents and massive gridlock in major urban areas. A power blackout covers the entire nation, and natural gas is not flowing, leaving millions in the cold. The economic system is completely frozen due to the elimination of financial data on central computers, and ATMs will not function. The networks of the Department of Defense, both classified and unclassified, have crashed, leaving the military a set of isolated units. Thousands would have died in the space of a few hours, and many more would do so in the days ahead as the effect of food and power shortages take their toll.

Clearly such a deadly cascade of effects from cyberattacks should be counted as an act of war. What is needed to recognize this reality is a focus not only on the means by which an attack achieves its effects, such as whether the deed is done by bombs or by computers, but also on the effects themselves. The effects of an attack play a significant role in determining whether the attack should be treated as an act of war, making just war theory and international law relevant to its assessment. In a study of cyberattacks, the National Research Council noted that the application of the terms force and armed attack "should be judged primarily by the effects of an action rather than its modality" (National Research Council 2009, p. 3). But the means or modality by which the effects are achieved should not be completely ignored. Michael Schmitt suggests the importance of appealing to consequences, but he rejects an exclusive reliance on consequences to determine what counts as an act of war. For example, he points out that economic and political coercion can have many of the negative effects of acts of war, though they are not treated by international law as acts of war (Schmitt 1998–1999, p. 908; National Research Council 2009, p. 257). For example, the economic sanctions on Iraq in the 1990s, on one estimate, led to the deaths of 239,000 children under five (Powel 1998). But there was no war in a legal sense waged against Iraq during most of the 1990s. The lethality of economic sanctions is distinct from the lethality of armed force, independent of the magnitude of the consequences. The question is on which side of this distinction the lethal effects of cyberattacks belong. Are they more like the effects of economic sanctions or more like the effects of armed force?

Clearly not all cyberattacks would count as acts of war. Cyberattacks cover a wide range of types and degrees of intrusion, and many of them are not even potentially lethal. In terms of types of attacks, some are passive and some are active. The passive intrusions may be intended simply to collect information (as in the case of an espionage attack, mentioned by Thomas Rid), while the active attacks are intended to affect or damage a computer system (and thereby often do damage in the physical world). Active intrusions can range from seeking to gain access in order to control a computer system, to implanting computer viruses or worms to destroy or corrupt data, to planting a "logic bomb" that is intended to lie in wait in a system ready to "explode" and do damage upon an internal or an external signal[5] (Schmitt 2002, p. 367). But, more to the point, the active intrusions can be intended or can achieve different degrees of physical damage. Schmitt claims that cyber attacks may or may not be acts of war, "depending on their nature or likely consequences" (Schmitt 2002, p. 375).

In order to distinguish cyberattacks that are acts of war from those that are not without appealing exclusively to consequences, Michael Schmitt seeks to determine the proper extension for the term "armed force." He proposes a "consequence-based interpretation" of the term. He claims that "the reference to armed forces is more

[5] Schmitt, "Wired Warfare," p. 367. The distinction between active and passive intrusions may be represented by the contrast between the Stuxnet worm (June 2010), which sought to damage nuclear centrifuges in Iran and the Flame virus (May 2012), apparently meant simply to collect information.

logically understood as a form of prescriptive shorthand for activity of a particular nature and intensity" (Schmitt 2002, p. 371, 396). The prescriptive shorthand implicitly takes into account not only the human suffering caused by an attack, but also the severity, immediacy, directness, and invasiveness of that harm. The use of armed force tends to have these characteristics to a high degree, while the use of economic and political sanctions does not, despite the fact that both may cause a great deal of human suffering. "Thus, the consequences of armed coercion are presumptively impermissible, whereas those of other coercive acts are not (as a very generalized rule)"[6] (Schmitt 1998–1999, pp. 914–915). This is the basis, in his view, of the distinction between harm imposed on one state by another that should count as an act of war from such harm that should not. Because the harmful effects of cyberattacks may be devastating, immediate, direct, and invasive, as in the scenario from Clarke and Knake (2010) presented above, it follows that cyberattacks can sometimes be acts of war. While Thomas Rid is correct to claim than most cyberattacks (including most of those that have occurred to date) are not acts of war, he is wrong to conclude that cyberattacks cannot be acts of war and that there cannot be a cyberwar. Stand-along cyberattacks, specifically those that fall into his category of sabotage, may, if severe enough, be acts of war.

So a cyberattack can be an act of war, and a war (a cyberwar) may be composed exclusively of cyberattacks. One interesting question that Schmitt's analysis seems to leave open is whether a stand-alone *operational* cyberattack could be an act of war. The Clarke scenario is an example of a *strategic* cyberattack, that is, one directed against the economic and social foundations of a society, but an operational cyberattack would be directed against military computers, attacking military command and control.[7] It seems that a large-scale operational cyberattack should count as an act of war. In terms of conventional war, an operational attack is a paradigm act of war, whereas a strategic attack is, in a sense, aberrational. If a strategic cyberattack counts as an act of war, so should an operational cyberattack. But, Schmitt's analysis seems to preclude an operational attack, at least one involving little collateral damage, being an act of war, the reason being that he places human suffering at the center of his case that a cyberattack may be an act of war. Consider this apparent paradox: cyber technology[8] promises the possibility of a major operational attack being achieved with much less human suffering than a conventional operational attack. The goal of an operational attack is disruption of the opponent's military, which a cyberattack might achieve by damage to the relevant computer systems and little harm to humans,[9] while a conventional operational attack, even with highly

[6] Given the potential severity of economic sanctions, as the Iraq sanctions indicate, the way might be open to challenge this categorization by positing that economic sanctions can also sometimes be acts of war.

[7] On the idea of strategic and operational cyberattacks, see (Schmitt 2002, p. 366; Arquilla 1999, p 389).

[8] I use the term "cyber technology," henceforth, to refer to the use of such technology for military purposes.

[9] For a discussion of this sort of cyberattack, see (Bayles 2001, p. 50).

accurate munitions, would blow things up and kill people. Just war theory is concerned to keep human suffering at a low level. So, the better an operational cyberattack would look in terms of just war theory, the less Schmitt's analysis would view it as an act of war. The result would be, at the extreme, that no cyberattack Schmitt would count as an act of war could be just, a result we should not accept. So, whatever the implications of Schmitt's analysis, I will regard a major operational cyberattack, not only a major strategic attack, as an act of war. (In what follows, I will use the term "cyberattack," unless otherwise qualified, to refer to an attack with consequences sufficiently serious to count as an act of war[10]).

Assume that one state launches a major cyberattack, whether strategic or operational, against another state. Where might things go from there? The state attacked might respond with a cyberattack, and if the war continued through a series of cyber exchanges, the result would be a cyberwar. But it seems more likely that such a war would become a conventional war, that the exchanges would move at some point from the cyberspace of the initial attack into physical space. The war would involve real bombs as well as logic bombs. This likelihood is important in understanding the moral assessment of cyberattacks, to which the next section is devoted.

7.2 Section II

Cyberattacks that are acts of war, as well as cyberwar more generally, like their conventional counterparts, are subject to normative regulation. There are rules of war, and some of these rules are moral or ethical, specifically, the rules of just war theory. This section discusses the applicability of the rules of just war theory to cyberattacks. The main question is whether the moral rules that apply to war in general are adequate or relevant to the phenomenon of cyberconflict. Does the moral theory traditionally applied to war address the novel moral issues raised by cyber technology? Does just war theory provide practical guidance to the cyberwarrior? Through its long history, just war theory has had to weather many social, political, and technological changes in the nature of war, changes which threatened to make the theory irrelevant or inapplicable in practice. But the theory has endured these changes largely intact, remaining relevant through many revolutions in military affairs. Can the same be said in regard to the technological changes that have created the possibility of cyberattack?

First, consider cyberattacks from the *ad bellum* perspective *Jus ad bellum* is composed of the moral rules concerning the initiation of war. Can these rules make sense of our moral choices of initiating a war through cyberattacks? The *ad bellum*

[10] As a terminological point, it should be noted that *any* cyberattack in the context of a conventional war could be referred to as an act of war. The claim in this paper that only cyberattacks causing sufficient physical damage would be acts of war refers to *stand-alone* cyberattacks, cyberattacks outside the context of a general war. This latter includes cyberattacks that initiate a war, a war which might then either continue as a cyberwar or become a conventional war, for example, if there was a conventional retaliation to the initial cyberattack.

rules are usually represented by a series of six criteria that must be satisfied before it is justified to go to war. A state is justified in initiating a war only if the war (1) has a *just cause*, (2) is declared by a *legitimate authority*, (3) is begun with a *rightful intention*, (4) shows *proportionality* between means and ends, (5) has a *reasonable chance of success*, and (6) is a *last resort*. Can these criteria represent an adequate standard by which to judge cyberattacks and cyberwar, as they do in regard to conventional war?

Some of the criteria are clearly problematic in the cyber context. Consider (1) just cause, often considered the most important of the *ad bellum* criteria. The paradigm just cause for a state's going to war is the opponent's aggression. A state is justified in using armed force when this is in defense against an act of aggression. The initial act of war may be a conventional attack or a cyberattack. But two problems arise in determining whether a state has a just cause when the initial attack to which it responds is a cyberattack. First, there is a *threshold problem*, and second there is the *attribution problem*. The threshold problem is that the initial attack must be an act of war, as opposed to a lesser act of force. If an attack using cyber technology falls short of an act of war, the state under attack has no just cause to go to war. The problem is exacerbated in the case of cyberattack due to the difficulties discussed earlier about distinguishing cyberattacks that are acts of war from those that are not. But this is a problem with conventional attacks as well; for example, a few shots fired across a border would not normally be an act of war and would not be a just cause for going to war. So the threshold difficulty is not a problem new to cyber technology.

The attribution problem is the difficulty of determining the source of an attack (Dipert 2010, p. 385, 401). Given the nature of cyber technology, it is often difficult to determine where at attack has come from and to justify this to the world, especially when the attacker seeks to hide its identity (Rowe 2010). This raises epistemological questions about the degree of certainty a state must have about who its attacker is before it has a just cause to respond with force (Dipert 2010, p. 393). In the case of conventional attacks, by contrast, the source of a major attack is usually obvious. John Arquilla claims, however, that the attribution problem in the case of cyberattack "is indeed difficult but is not insurmountable." He notes that attribution is sometimes a problem in the case of conventional attacks as well, for example in the historical case of the source of "phantom" submarine attacks on merchant ships bringing supplies to the Loyalists in the Spanish Civil War. He shows how this case was effectively dealt with (Arquilla 1999, p. 396). It may arise as well in the case of conventional terrorist attacks by state-sponsored agents (or in an earlier era, by pirates). Arquilla argues that one can make inferences based on the purposive nature of acts of war, along with other detection techniques, allowing a state usually to discover with a sufficient degree of certainty the source of anonymous attacks. We will revisit the attribution issue in the next section.

But there is another dimension to the attribution problem, namely, the disruptive cyber activity of independent individuals commonly referred to as hackers. Arquilla speaks of the way "in which the information revolution empowers small groups and individuals to wage information warfare" (Arquilla 1999, p. 394). The

means of cyberattack are inexpensive and widely distributed, so that anyone with the proper skills can engage in disruptive or damaging activities over the internet. Arquilla himself raises the problem of hackers not in connection with the attribution problem, but instead in connection with the criterion of (2) legitimate authority. This criterion requires that anyone engaging in war have the authority within a large political organization such as a state to do so. Hackers obviously do not have legitimate authority in regards to acts of war. But the existence of hackers does not pose a problem for legitimate authority. *Ad bellum* rules require that those making war be legitimate authorities. But hackers are not making war, since war is a conflict between large political organizations with lines of authority. That hackers may cause a lot of damage in not a problem for the rules of war. It is rather a problem of law enforcement (as is the problem of pirates when they are not state-sponsored).

But hackers may represent a dimension of the attribution problem for the criterion of just cause. If a state finds itself under cyberattack, and if the attack might have come from hackers rather than from a state, the problem of attributing the attack to a particular state for the sake of establishing a just cause for a military response is obviously more difficult. But how much does the hacker phenomenon exacerbate the attribution problem? There are two views on this. One is connected with the common perception that cyberattacks are a weapon of the weak, an equalizer between the weak and the strong, whether the weak happen to be a state, a small independent group, or an individual hacker.[11] On this view, the relatively powerless, including hackers, can through cyberattack do outsized damage to powerful states. Joseph Nye endorses this view. He offers a supporting quotation from a US military official ("Sooner or later, terror groups will achieve cyber-sophistication.") and cites another who argues that "while states have the greatest capabilities, nonstate actors are more likely to initiate a catastrophic attack" (Nye 2011, pp. 21–22).

Another view is that the production of effective cyber weapons is an "expensive, skilled, labor-intensive [and] state-centric enterprise" (Lucas Cyberwar, p. 18). Hackers can be disruptive, shutting down websites and such, but cannot do the high level of physical damage that would be equivalent to an act of war, whether strategic or operational. As evidence for this perspective, one could cite the Stuxnet worm, designed to interfere with the centrifuges Iran was using to refine uranium. The widespread view is that the complexity and sophistication of Stuxnet required that it be produced with the resources of an advanced state[12] (Lucas Cyberwar, pp. 14–16). Even so, might the hackers catch up over time, as Nye suggests? Perhaps, but they would be aiming at a moving target. Strong states will be developing their defensive as well as their offensive capabilities, and any increase in offensive capability by the hackers may be more than compensated for by their targets' increase in defensive capability. If this second view is correct, then the activities of non-state hackers (or, to a lesser extent, weak states) do not add greatly to the attribution problem. A sophisticated cyberattack will reveal the hand of a powerful state. States may seek to

[11] This view is expressed, for example, in (Schmitt 1998–1999, p. 897).

[12] This view was confirmed by a news article documenting how Stuxnet was a project of the United States and Israel (Sanger 2012).

use apparently independent agents, so called patriotic hackers, to mask their identity (Arquilla 1999, p. 387). But a (relatively) sophisticated attack could justifiability be attributed to a state, whether the attack came directly from the state or from apparently independent agents the state is using to cloak its involvement. If an attack is (relatively) unsophisticated, it may be assumed to come from a non-state source, in which case seeking out the hackers to hold responsible would be a matter of criminal law (Bayles 2001, p. 55).

What about the other *ad bellum* criteria? Criterion (3) rightful intention requires that a war be initiated with the intention to address the just cause for the war, and this seems to apply to cyberattacks in a straightforward way. In addition, criterion (5) reasonable chance of success, which requires that a war not be a hopeless cause, also seems to apply unproblematically to wars initiated by cyberattack. Criterion (4) proportionality requires that a war be reasonably expected to produce more good and harm. Its application to cyberattack may not be so clear. On the one hand, proportionality seems potentially more easily satisfied by a cyber war, given that cyberattacks are in general easier to carry out with a minimal loss of life than conventional attacks. On the other hand, there is a factor that militates against this. In general, cyberweapons cannot be tested, because to test them may be to reveal to the opponent how it needs to adjust its systems to defend against that mode of attack, for example, what antivirus patch it needs to develop. For this reason and others, it may be unusually difficult to predict how effective a cyberattack would be. As a result, states may tend to err on the side of a larger response, which would make proportionality more difficult to satisfy (Rowe 2010).

The most serious problem posed by cyberattack to the *ad bellum* rules may be arise in the case of the criterion (6) last resort. This criterion requires that a state go to war only if it has no reasonable peaceful alternative. This criterion is crucial to the success of just war theory in limiting the occurrences of war. There may often be cases where a state has a just cause to go to war (and where other criteria are satisfied as well), but where there are peaceful alternatives that may resolve the conflict. The purpose of the last resort criterion is to insure that war is not resorted to in such cases, at least until peaceful alternatives have been shown to fail. Arquilla argues that this criterion is one of the respects in which cyberattack technology plays havoc with the traditional morality of war, leaving "just war theory in tatters" (Arquilla 1999, p. 394). The main way in which cyber technology undermines the applicability of the last resort criterion is in the tendency of the technology to encourage *anticipatory war* (preventive or preemptive war), which is war initiated to avoid a perceived future threat from one's opponent. For states with the requisite cyber technology, it may seem so easy and tempting to initiate war in a conflict situation that the result would be that the requirement of last resort is effectively ignored. There are several reasons for this. First, an operational cyberattack may seem like such an obvious thing to undertake when a state perceives a future threat from its opponent, given that such an attack promises severe disruption of the opponent's military capability without a great deal of physical destruction. Second, such a disruption provides the attacking state with a great military advantage, perhaps an effective decapitation by itself forcing the opponent's surrender. Third, a state

may believe that its operational cyberattack, because it would cause little physical damage, would not even be considered an act of war. I have argued that this is an illusion, given that severe military disruption must count as an act of war, but it may be an illusion to which states are prone. Moreover, even if an initial cyberattack was short of an act of war, the likelihood of escalation to war would be very great. I will return to this issue in the next section.

Now consider cyberattacks in the context of the other aspect of just war theory, *jus in bello*, the morality of how a war is fought. Again, there are a set of criteria, in this case *discrimination*, *proportionality*, and *due care*. Discrimination requires that attacks in war be directed against military targets rather than civilian targets. Proportionality requires that attacks of war be such that the contribution they are expected to make to victory in the war outweighs the expected amount of harm they would do. (*In bello* proportionality differs from *ad bellum* proportionality in that it is applies to individual military actions rather than to the war as a whole and does not assume a just cause). Due care requires that attacks in war minimize expected harm to civilians. Some argue that it would be easier to keep cyberwar within the limits defined by these criteria than to keep conventional war so limited, that cyberattacks "if rightly handled, could end up being more discriminate, more proportional, and thus more in compliance with… the moral principles of *jus in bello*, than any conventional counterpart" (Lucas 2010, p. 297). But this judgment is hasty. Cyberattacks raise some problems with each of these criteria.

Consider discrimination. This criterion would, of course, rule out strategic cyberattacks, as it rules out strategic conventional attacks. Military objects can be deliberately attacked, but civilian objects cannot. Civilian infrastructure cannot be made the object of military attack. But can a clear line be drawn between military and civilian objects? This raises the problem of so-called dual-use infrastructure, infrastructure that serves both military and civilian purposes, such as electrical power grids. This is also a problem in the case of conventional attack. For example, the United States has taken a permissive view on what dual-use infrastructure it may attack. In the first Gulf War, it treated the electric power grid of Iraq as liable to attack. But these attacks resulted in the deaths of an estimated seventy to ninety thousand civilians (Bayles 2001, p. 52). In the indirect deaths of the civilians were taken into account, this seems like a strategic attack rather than an operational attack (though the issue of intentionality would complicate this judgment).[13] Although this problem arises in conventional war (as it did in the Gulf War), it is a special problem in the case of cyberattack because infrastructure is a natural point of attack in cyberspace (Hirschland 2001, p. 11). Infrastructure is to an increasing extent under the control of computer systems.

This dual-use problem is exacerbated in the case of cyberattack because such attack seems benign in comparison with a conventional attack. Destroying the electrical grid of a nation with conventional weapons, even precise ones, would likely kill hundreds of civilian power workers, while to do so with a cyberattack may kill

[13] For an argument against an understanding of the rules of war that would allow such a permissive view of the liability of dual-use infrastructure to attack, see (Shue and Wippman 2002).

no one directly. This is part of the illusion that cyberattack is a bloodless strategy, which was discussed in the first section. It is an illusion, as the facts of the Gulf War illustrate. One way to think of the illusion is this. Consider that the purpose of traditional kinetic strikes against enemy combatants is not directly to kill them, but rather to disable them, to make them unable to resist one's own forces (this is the basis of the rule of war protecting injured combatants from attack). It just so happens that with present technology the only effective way to disable combatants is usually to kill them. The illusion is that cyber technology seems to promise a way to disable the opponent as a whole by destroying infrastructure without killing anyone. It is an illusion because the destruction of the infrastructure will lead to large numbers of civilian deaths.

In addition, there are two other special problems for the criterion of discrimination posed by cyberattacks. First is the problem of the combatant status crucial to the application of discrimination. The criterion assumes that there is a clear distinction between combatants and civilians, but cyber technology muddies the distinction due to "the use of typically civilian technology and know-how to conduct military operations via computer" (Schmitt 2002, p. 398). Discrimination becomes more difficult to apply because many civilians will be intimately involved in the activity of war. Second there is the problem of *perfidy*. Deception is a recognized part of war, and most deception, referred to as *ruse*, is permissible, but some deception, perfidy, is not acceptable. One example of perfidy is the feigning of a status protected under the rules of war, such as combatants pretending to be civilians. This is a violation of discrimination. Cyber technology offers great opportunity for deception in general and perfidy in particular. For example, a state could, in an act of perfidy, plant an "all clear" message into the opponent's communications systems just before an attack (Bayles 2001, p. 50). Cyber technology would increase the opportunity and temptation for states to engage in perfidy.

The *in bello* criterion of proportionality works in tandem with the criterion of discrimination under the moral framework know as the *doctrine of double effect*. The idea is that while discrimination precludes attacks intended to harm civilians, some expected civilian harm, if not intended, may be permissible, just in case it satisfies proportionality. This sort of moral calculation arises especially in the case of cyber technology because, as mentioned, the natural targets of cyberattack are the computer systems controlling the infrastructure on which civilians depend on for survival. Here the illusion that cyberattacks are bloodless again plays a role, leading those applying the doctrine of double effect to tend to ignore the long-term harm to civilians from infrastructure attacks. In addition, there is the problem mentioned earlier in discussion of *ad bellum* proportionality that the uncertain expectations about the effects of a cyberattack could lead the attackers to launch a more devastating attack to insure that it has the desired effects. Added to this is the potential for what are called "reverberating effects," which is the tendency, due to interconnectivity, for effects in one realm or region to produce effects in another, often in a completely unpredictable way (Schmitt 1998–1999, pp. 893–894). All of these points show how the criterion of proportionality would be more difficult to satisfy in the case of cyberattacks.

But as a counterweight to these concerns, there is, also as mentioned earlier, a way in which the criterion of proportionality may be more easily satisfied through the use of cyber technology. While cyberattacks can easily impose great costs on civilians, even when these costs are not intended or even foreseen, they also may potentially be used in a way that imposes lesser civilian costs than corresponding conventional attacks would do. The difference lies in the way cyberattacks and conventional attacks do the damage they do. In the case of an opponent's electrical power grid, for example, destruction through conventional attacks would mean that the grid would be out of commission for weeks or months, if not longer, given the need to rebuild it. But a cyberattack could probably be designed to knock out the grid only temporarily, given that it could be done without any destruction of the facilities. The destruction might be done in a way that was effectively reversible (National Research Council 2009, p. 264). Indeed, cyberattacks "may make it possible to achieve desired military aims with less collateral damage and incidental injury than in traditional kinetic attacks" (Schmitt 2002, p. 397).

This potential difference in the minimal amount of destruction with which cyberattack and conventional attack can be carried out has an important bearing on the relevance of cyber technology to the third *in bello* criterion, due care. Due care requires that an attack be done in the way that minimizes the amount of civilian harm, and for a state with the relevant technological capability, this will usually be through a cyberattack. Michael Schmitt suggests that "military commanders will in certain cases be obligated to employ their cyber assets in lieu of kinetic weapons when collateral and incidental effects can be limited" (Schmitt 2002, pp. 397–398). Following the demands of the due care criterion, the acquisition of cyber capability and its use in preference to conventional attack may become morally obligatory!

In summary, there are some difficulties (and also some advantages) that cyber technology potentially poses for the application of the criteria of *jus in bello*. But the difficulties seem not to be sufficient to find that the technology threatens to make just war theory irrelevant. Many of these are connected with the illusion of bloodlessness, and this is something that combatants and military leaders can be educated to reject. Cyber war is not a new kind of war, in the sense that it requires different moral rules about how it is fought. A similar judgment seems appropriate for the criteria of *jus ad bellum*, with one important exception. For all of the *ad bellum* criteria save one, the difficulties we have considered that arise when they are applied to cyberattacks are not sufficient to find that the technology threatens to make just war theory irrelevant. The one exception is the criterion of last resort. A case could be made that cyber technology would make this criterion inapplicable, at least in practice, threatening the relevance of just war theory to cyberattack. Whether this is in fact the case is considered in the next section.[14]

[14] Of course, I must note that the judgments made in this section, as with other of the judgments in this essay, are at least partly speculative, given that the future development of the technology and the way it turns out to be applied in the real world of war cannot be accurately predicted.

7.3 Section III

Speaking of the implications of the military use of cyber technology, Arquilla notes that "the one area that may change is the use of force in preventive ways" (Arquilla 1999, p. 387). The greatest problem that cyber technology poses for just war theory is the potential that this technology could lead to a great increase in anticipatory wars, thereby vitiating the likelihood that war initiation through cyberattack could satisfy the *ad bellum* criterion of last resort. Anticipatory war is a war where the belligerent strikes the first blow while justifying the attack on the defensive grounds that the attack is in response to an expected attack from the opponent. Among anticipatory wars are preventive wars and preemptive wars. Preventive wars occur when the attacker believes that its opponent intends to strike at some indefinite time in the future. Preemptive wars occur when the attacker strikes first in reasonable fear that it opponent's own first strike is imminent. In both sorts of case, the attacker believes that if it lands the first blow it is more likely to win the war it expects sooner or later. Preventive wars always and preemptive wars sometimes violate the criterion of last resort because they ignore the peaceful alternatives that may be available to avoid war.

In order to draw some conclusions about the relevance of last resort to cyber conflict, we need to make a brief excursion into strategic thought as it applies to cyberattack. One of the reasons that anticipatory war is a special problem in the context of cyber technology is that cyberattacks are potentially very effective in an effort to "prepare the battle space," that is, to weaken an opponent in preparation for a conventional attack. Preparatory cyberattacks could provide a great advantage in a conventional war by disrupting the opponent's military communications, its intelligence gathering assets, its global positioning systems, and so forth (Schmitt 1998–1999, p. 929). The capacity of cyber technology to achieve such results is one reason why with this technology, as with nuclear weapons technology, the offense dominates the defense (Nye 2011, p. 21). It is easier to destroy assets with cyberattack than to protect them from cyberattack. All of these factors lead to a situation of *crisis instability*, a situation in which war is more likely to break out in a crisis because both sides have incentives to initiate an attack (National Research Council 2009, p. 306). The upshot is that cyber technology "makes war more thinkable." (Arquilla 1999, p. 398). George Lucas raises the concern that cyber technology will "lower the threshold for resorting to war of any sort, traditionally consigned to being the last (rather than the earliest) resort to conflict resolution with adversaries or competitors" (Lucas 2010, p. 294). War becomes more thinkable, and the last resort criterion is devalued or ignored.

These points may be developed by our considering some comparisons between the strategic implications of cyber and nuclear technology. While the two technologies differ dramatically in the amount of destruction they can cause, there are comparisons in the strategic environment each creates[15] (National Research Council

[15] Joseph Nye notes that a strategic cyberattack could send the economy back to 1990, while a strategic nuclear attack could sent the economy back to the Stone Age (Nye 2011, p. 22).

2009, p. 295). Both represent a dominance of the offense over the defense. Consider first how offense dominance works in case of nuclear technology. First, compare conventional deterrence and nuclear deterrence. In the case of conventional deterrence, a state's effort to deter its opponent is based on the threat both to inflict costs and to deny benefits, on a threat of *punishment* and on a threat of *denial*. Denial is the ability of a state's defensive capabilities to blunt the success of an attack. So, even if deterrence fails, the success of an attack is not guaranteed. But in the case of nuclear weapons, defenses are ineffective because there is no adequate way to stop a large number of nuclear warheads on missiles from getting through to their targets. There is no denial and nuclear deterrence rests exclusively on the threat of retaliatory punishment. It might seem that this would lead to great crisis instability, as there would be great advantage to going first, but this turned out not to be the case. The reason is that both sides were able to proliferate and protect warheads, making their retaliatory capacity partly invulnerable to surprise attack; each side was guaranteed to have enough warheads left over after a surprise attack to destroy the attacker. Each side had the capacity for assured destruction, and together the United States and the Soviet Union were in a state of MAD, mutual assured destruction. There was no advantage and so no incentives for going first because whichever side went first, both sides would be destroyed.

Now consider some analogies (and disanalogies) between these features of nuclear strategy and the potential strategic environment of cyber technology.[16] Consider first offense dominance. This does not mean quite the same thing for cyber technology as it does for nuclear technology. In the case of nuclear technology offense dominance is due to the destructive power of the offense and the impossibility of effective defense. In the case of cyber technology offense is not as destructive and defenses may have some effectiveness. On the side of defenses, offense dominance is due to uncertainty about how effective cyber defenses would be and to the greater cost of defense as compared with offense. As Randall Dipert notes, cyber defense is unlikely to be sufficiently successful and likely to be too expensive (Dipert 2010, p. 403). On the side of offenses, offense dominance is due to disruptive effects of operational cyberattacks and the social destruction possible with strategic cyberattacks. But the social destruction of cyberattack is much more tolerable, much less of a punishment, than that of nuclear attack. What is more to the point, the advantage of operational cyberattack comes mainly from going first, creating incentives to strike first that may not be outweighed by the threat of punishment. In the case of nuclear deterrence, in contrast, the threat of punishment far outweighs the advantages from going first.

Consider other features of cyber deterrence. (I understand cyber deterrence to mean deterrence *of* a cyber attack, whether *by* cyber threats or conventional threats). There are some features of cyberattack that make cyber deterrence less credible, hence less effective. The most important of these is the attribution problem, dis-

[16] This discussion largely concerns adversarial relations between "near-peer" states, those roughly equal in military capability. Different factors may arise in the relations between adversaries in an asymmetrical power relationship.

cussed earlier. A state that believes that it can attack anonymously because its attack will not be successfully attributed to it will not be deterred by a threat of retaliation. Even if attribution can be achieved in most cases, it will likely take time, and threats of delayed retaliation are less credible than threats of immediate retaliation. In contrast with the nuclear threat, "the difficulties of attack attribution leave a comparable [cyber] threat with far less credibility"[17] (National Research Council 2009, p. 2, 294, 295). At the same time, the fact that cyber defenses could have some effectiveness means that cyber deterrence (unlike nuclear deterrence) has an element of denial, increasing, to that extent, its credibility (Nye 2011, pp. 33–34).

Using cyber threats to deter cyberattacks may not be an effective form of deterrence due, among other reasons, to uncertainty about how effective cyber retaliation would be. (Again, in contrast, there is little uncertainty about the effects of a nuclear retaliation). For this reason, conventional threats (or even nuclear threats) may be a necessary part of an effective posture of deterrence of cyberattacks. This seems to be current U.S. policy, as the Whitehouse has declared: "When warranted, the United States will respond to hostile acts in cyberspace [reserving] the right to use all necessary means" (Whitehouse 2011, p. 14). But including conventional threats to deter cyberattack would pose problems of its own. A conventional response to cyberattack, Arquilla remarks, "may tend toward escalation" (Arquilla 1999, p. 390). The reason is that such a response would be perceived as upping the ante. This perception may result from two features of the situation. First, the amount of harm done by the initial cyberattack may not be clear, not only to the attacker but to the victim, and second, a comparison between harm done in a cyber attack and harm done in a retaliatory conventional attack is inherently difficult to make (Arquilla 1999, p. 391). Given the bias each side has toward its own case, these two features could lead to a perception on the part of the recipient of the retaliatory attack that that attacker has upped the ante, thereby calling for the recipient to up the ante further in response. This means that throwing conventional attacks into a military exchange begun with cyberattacks would make the *signaling* necessary to avoid escalation more difficult (National Research Council 2009, p. 308). An escalatory spiral could easily result. This sort of situation suggests a systemic weakness of cyber deterrence. The situation has a dilemmatic structure: cyber deterrence can be restricted to cyber threats or can include conventional threats as well; if the former, the deterrence posture is weak, and if the latter it is weak as well.

So cyber deterrence is weak due to the attribution problem, and it is weak due to the sort of dilemmatic structure just noted. When deterrence is weak, the likelihood of one side or the other initiating a cyberattack is greater. This contributes to instability, where the likelihood of one side initiating a cyberattack rises even further. If side A recognizes that side B is more likely to initiate a cyberattack (because deterrence is recognized to be weak), side A itself becomes more likely to initiate a

[17] Another factor in the need for a delay before the retaliatory response is that it may take time for the victim of a cyberattack to figure out how much damage was done, which it needs to know before it can decide how great the retaliation should be (or even whether it should occur at all) (National Research Council 2009, p. 310).

cyberattack out of fear that B might do so, which then leads to B being more likely to do so, and so forth. This dynamic has been called in the case of nuclear strategy the reciprocal fear of surprise attack. Surprise attack would be an anticipatory war. This creates crisis instability because a crisis in the relationship between A and B is the time when this dynamic is most likely to engage. In the nuclear context, this dynamic is forestalled by the prospects of mutual destruction, but this prospect is not available to forestall the dynamic in the case of cyberattack.

Now we may return from our excursion into cyber strategy to the discussion of *jus ad bellum* and the criterion of last resort. Anticipatory war is generally a violation of last resort, always in the case of preventive war and often in the case of preemptive war. The maturation of cyber military technology will increase the risk of anticipatory war, due to the weakness of cyber deterrence, along with the tendency of potential belligerents to treat a cyberattack as less than a full-fledged act of war, compounded by the fact that there is a deep inherent advantage to going first, due in part to the way in which initial cyberattack works to "prepare the battle space" for a more general war. Due to crisis instability, the pressure on states to initiate cyberattack will sometimes be great, and this pressure means that the last resort criterion will often be ignored. War becomes more thinkable because the last resort criterion is not being thought about.

The earlier arguments have shown that, in prospect, cyber technology make the other *ad bellum* criteria and the *in bello* criteria more difficult, but not impossible to adhere to. But the dynamics of cyberattack and cyber deterrence may show that the last resort criterion is, as a matter of practice, impossible to adhere to. The fear that each side has, especially in a crisis, that the other is about to attack will make it often impossible for either side to effectively explore options short of war for resolving the conflict. In this sense, cyber military technology makes this criterion of last resort irrelevant, and to this extent the maturation of cyber military technology would take us beyond just war theory.

References

Arquilla, John. 1999. Ethics and information in warfare. In *The changing role of information in warfare,* ed. Z. Khalilzad et al. 379–401. Santa Monica: Rand Corporation.

Bayles, William. 2001. The ethics of computer network attack. *Parameters* 31:44–58.

Clarke, Richard, and Robert Knake. 2010. *Cyber war: The next threat to national security and what to do about it*. New York: Harper Collins.

Dilpert, Randall. 2010. The ethics of cyberwarfare. *The Journal of Military Ethics* 9:384–410.

Hirschland, Matthew. 2001. Information warfare and the new challenges to waging just war presented at conference of the American Political Science Association, Denver, CO.

Lucas, George R. 2011. Permissible preventive cyber warfare, Proceedings of the Air Force Research Institute on the Future of Cyber Power, ed. Pano Yanakageorgos, et al.

Lucas, George R. 2010. Postmodern war. *Journal of Military Ethics* 9:289–298.

National Research Council. 2009. *Technology, policy, law, and ethics regarding U.S. acquisition and use of cyberattack capabilities*. Washington, D.C.: National Academies Press.

Nye, Joseph. 2011. Nuclear lessons for cybers. *Strategic Studies Quarterly* 5 (4):18–38.

Powell, Michael. 1998. The deaths he cannot sanction; Ex-U.N. Worker details harm to Iraqi children. *Washington Post*, 17 December.

Rid, Thomas. 2011. Cyber war will not take place. *Journal of Strategic Studies* 35:5–32.

Rowe, Neil. 2010. The ethics of cyberweapons in warfare. *International Journal of Cyberethics* 1:20–31.

Sanger, David. 2012. Obama order sped up wave of cyberattacks against Iran. *New York Times*, 1 June.

Schmitt, Michael. 1998–1999. Computer network attack and the use of force in international law: Thoughts on a normative framework. *Columbia Journal of Transnational Law* 37:885–937.

Schmitt, Michael. 2002. Wired warfare: Computer network attack and *Jus in Bello*. *International Review of the Red Cross* 84:365–398.

Shue, Henry, and David Wippman. 2002. Limiting attacks on dual-use facilities performing indispensable civilian functions. *Cornell International Law Journal* 35:559–577.

Whitehouse. 2011. International strategy for cyberspace. http://www.whitehouse.gov/sites/default/files/rss_viewer/international_strategy_for_cyberspace.pdf.