JOURNAL OF CYBERSECURITY

Journal of Cybersecurity, 1(1), 2015, 53–67
doi: 10.1093/cybsec/tyv003
Advance Access Publication Date: 28 November 2015
Research Article

Research Article

Tipping the scales: the attribution problem and the feasibility of deterrence against cyberattack

Jon R. Lindsay*

Munk School of Global Affairs, University of Toronto

*Corresponding author: E-mail: jon.lindsay@utoronto.ca

Received 18 May 2015; revised 22 September 2015; accepted 28 September 2015

Abstract

Cyber attackers rely on deception to exploit vulnerabilities and obfuscate their identity, which makes many pessimistic about cyber deterrence. The attribution problem appears to make retaliatory punishment, contrasted with defensive denial, particularly ineffective. Yet observable deterrence failures against targets of lower value tell us little about the ability to deter attacks against higher value targets, where defenders may be more willing and able to pay the costs of attribution and punishment. Counterintuitively, costs of attribution and response may decline with scale. Reliance on deception is a double-edged sword that provides some advantages to the attacker but undermines offensive coercion and creates risks for ambitious intruders. Many of the properties of cybersecurity assumed to be determined by technology, such as the advantage of offense over defense, the difficulty of attribution, and the inefficacy of deterrence, are in fact consequences of political factors like the value of the target and the scale-dependent costs of exploitation and retaliation. Assumptions about attribution can be incorporated into traditional international relations concepts of uncertainty and credibility, even as attribution involves uncertainty about the identity of the opponent, not just interests and capabilities. This article uses a formal model to explain why there are many low-value anonymous attacks but few high-value ones, showing how different assumptions about the scaling of exploitation and retaliation costs lead to different degrees of coverage and effectiveness for deterrence by denial and punishment. Deterrence works where it is needed most, yet it usually fails everywhere else.

Key words: strategy; deterrence; policy; cybersecurity; information security; international relations.

Introduction

Is it possible to prevent attacks on vital information systems? The global epidemic of cybercrime, embarrassing breaches of firms like JPMorgan Chase and Target, relentless Chinese industrial espionage, Edward Snowden's revelations about National Security Agency surveillance, and the Stuxnet disruption of Iranian enrichment all offer ample reasons for pessimism. Network complexity, easy anonymity, and user gullibility enable attackers to exploit vulnerabilities faster than defenders can patch them. The recent US government indictments of Chinese spies that will never be prosecuted, minor sanctions on an already isolated North Korea for hacking Sony, dithering in the wake of the Office of Personnel Management penetration, and advocacy for voluntary and unenforceable norms, moreover, undermine confidence that serious attacks will ever be adequately punished even if attribution is successful. The absence to date of any "digital Pearl Harbor" or "cyber 9/11" catastrophes provides cold comfort.

Scholars and policy analysts are generally pessimistic about cyber deterrence, noting considerable confusion about what the concept even means [1–3]. Deterrence theory distinguishes deterrence by denial, the threat that effective defenses will defeat an attack, from deterrence by punishment, the threat that costly retaliation will offset the benefits of a successful attack [4]. Cyberspace seems to undermine both strategies because offense is relatively easier than defense [5], and the attribution problem precludes reprisal, i.e. punishment seems to require a return address [6, 7]. Furthermore, because cyber weapons depend on "zero day" vulnerabilities that will likely be countered if publicly revealed, it is hard to credibly threaten to use them for retaliation or other coercion [8]. Faced with a choice between bad alternatives, many argue that denial should be

emphasized over punishment [9, 10]. Policy should thus focus on creating defense in depth via intrusion detection and counterintelligence [11], improving user "hygiene" [12], and building resilient systems that can tolerate the inevitable attacks [13]. As a recent US Deputy Secretary of Defense wrote, "deterrence will necessarily be based more on denying any benefit to attackers than on imposing costs through retaliation. The challenge is to make the defenses effective enough to deny an adversary the benefit of an attack despite the strength of offensive tools in cyberspace" [14].

These arguments are not wrong, but they are incomplete. The technical challenges of attribution depend on critical organizational and political context that is making attribution, and thus deterrence, harder and easier along different dimensions [15]. Attackers depend on deception to obfuscate their identity, but they are more likely to make mistakes against complex targets while defenders are more likely to use deception themselves to protect the targets they most value [16]. Indeed, cyber operations are unsuited for coercive signaling for the same reasons that they are useful for marginal adjustments in the distribution of power, at modest scales where offensive deception is more likely to work. Furthermore, there is no reason why retaliation for cyber attacks should be confined to the cyber domain, especially in response to particularly egregious acts, just as the USA saw no reason to retaliate for 9/11 by flying airplanes into buildings in Kabul. Attribution and deterrence become feasible where defenders most need them to work, even as everyday failure is ubiquitous below the threshold of credible retaliation.

Given that most attacks tend to fall on the lower end of the value spectrum, conducted for criminal gain, surveillance, or protest rather than physical combat [17, 18], it is unsurprising that many are skeptical of deterrence. Cheap talk – unenforceable indictments, minor sanctions, and "name and shame" tactics – has not prevented historical attacks, and policymakers perceive more significant punishment to be too costly, too risky, or too escalatory to enforce. Yet deterrence failure here tells us little about the ability to deter attacks against higher value targets where the costs of attribution and retaliation relative to target value may be much less. The most significant attack on critical infrastructure to date, the Olympic Games operation against Iran, is notable for the restraint exercised by the actors, evidenced by an apparent goal of marginal disruption rather than catastrophic destruction, years of meticulous planning and surveillance, extensive efforts to avoid compromise and minimize collateral damage, and the significant power advantage of the USA over Iran to deter serious retaliation [19]. The absence of evidence of serious attacks might be interpreted as evidence that deterrence is working to protect some of the most valuable targets, or simply that sufficiently capable actors do not perceive a benefit in attacking [8] (Distinguishing between deterrence and disinterest in the cases of dogs that do not bark is a persistent difficulty in empirical deterrence research. This article focuses on the theoretical feasibility of deterrence at scale).

Cybersecurity is an inherently interdisciplinary problem because digital threats depend on both technology and incentives. The field of international relations (IRs) has a long tradition of understanding the sources of conflict and ways it can be avoided, so it is reasonable to leverage it to make sense of conflict in cyberspace [20–28]. Cybersecurity can also be a source of theoretical puzzles that can help to sharpen concepts within particular disciplines. Some scholars argue that cyberspace upends traditional theory altogether [29–32], but one need not go to such an extreme. The attribution problem, for instance, too often viewed through a technical lens with deterministic results for deterrence, can be grounded within IR theory to understand its implications for cyber conflict. Conversely, although the effects of uncertainty about capabilities and interests are well understood in IR theory, anonymous attacks involve uncertainty about the very identity of the opponent. Anonymity is certainly not unique to cybersecurity, but its prevalence raises questions about how strategic bargaining works, and fails, in the digital domain.

This article leverages rationalist IR theory to both explain and bound the pessimism about cyber deterrence and incorporate the newly salient problem of attribution. I find that different assumptions about the costs of attribution and retaliation relative to target value lead to different conclusions about the effectiveness and coverage of deterrence by denial and punishment. I proceed in four parts. I first examine the effects of information asymmetry and commitment problems in cybersecurity. Next I introduce the challenge of anonymity and argue that the attribution problem is scale dependent. The third section uses a simple formal model to show how the recalcitrance of attribution and the feasibility of deterrence depend on assumptions about costs and value. Few assumptions are needed to model the historical distribution of many low-value attacks and few to no high-value attacks. The concluding section discusses policy implications, arguing that toleration for low-level aggression is the price of credible deterrence against serious attacks.

Politics by cyber means

Most IR scholars assume that political actors use war and threats of war instrumentally in the pursuit of political goals such as power, security, wealth, and status (Mainstream realist and liberal approaches to IR share a rationalist assumption that actors compare costs and benefits when choosing a strategy. They differ mainly on the relative importance of the international balance of power versus domestic politics and institutions for structuring policymaker incentives. Some constructivist or postmodern approaches dispute the rationality assumption, instead emphasizing the role of ideas, culture, and interpretation in the social construction of belligerent policy. Because it is helpful to understand rational incentives for strategic behavior prior to assessing the relevance of departures, this article assumes that actors who employ cyber weapons tend to respond to incentives more or less rationally, although uncertainty is rife and actors make mistakes). As Clausewitz famously put it, war is "a true political instrument, a continuation of political intercourse, carried on with other means. What remains peculiar to war is simply the peculiar nature of its means . . . . The political object is the goal, war is the means of reaching it, and means can never be considered in isolation from their purpose" [33]. This insight is captured in modern theories that treat conflict as a process of rational bargaining over some contested good [34–37]. This section applies this paradigm to cyber conflict, highlighting factors that both promote the use of covert intrusions to marginally alter the distribution of power but limit their utility for making coercive threats or major power adjustments.

Information and commitment

The possibility of using force to alter the distribution of power can never be ruled out in anarchy, where there is no overarching political authority to adjudicate disputes and enforce agreements [38]. Even so, the outbreak of war is something of a puzzle if one assumes fighting is costly (If war were attractive for its own sake there would be little question about why wars occur. Historically, some cultures have imbued war with positive ritualistic value [39]). Why should both sides decide to shed blood, waste treasure, and destroy some of the disputed resource through fighting when they could each be better off negotiating to revise or preserve the status quo in favor of stronger actor? Two important answers are that information

asymmetry and commitment problems lead to inefficient bargaining outcomes [35]. Actors might want to avoid war, but they also want a favorable distributional outcome, so they are likely bluff, conceal capabilities, and break promises. These problems are legion in cyberspace.

Uncertainty about power and resolve is one of the most important factors contributing to the outbreak of war [40, 41]. One or both actors may not understand, or be unable to credibly communicate, the true balance of power, the value they each place on the dispute, or the likely outcome of a conflict. Because both understand that the other has incentives to misrepresent strength to get a better deal without fighting, they are inclined to discount threats that might, if believed, correct the misunderstanding. An actor whose power depends on secret plans or weapons, moreover, may be unable to make credible threats in a crisis because revelation would enable the opponent to develop countermeasures that nullify the advantage [42]. In September 1950, China could not advertise that it had infiltrated over 300 000 troops into North Korea, a fact that probably would have dissuaded US intervention. China warned the USA not to cross the 38th parallel, but American policymakers discounted the ambiguous threat and pushed north. Japan likewise could not coerce the USA in 1941 with the threat of a devastating attack on Pearl Harbor because the Pacific Fleet would simply have reoriented its defenses to defeat the raid. Japan furthermore misjudged the resolve of the USA to wage total war to avenge the insult. False optimism thus leads one or both actors to prefer conflict over negotiation [43, 44].

Even with perfect information, actors may be unable to credibly commit to a deal because they have incentives to renege in the future [45, 46, 47]. A declining state may prefer to fight a preventative war today rather than be forced to accept an undesirable peace tomorrow when the rising state will have more bargaining power. Leaders may be unable to restrain their bureaucratic agents, proxy forces, or angry citizens from pressing attacks after a formal peace deal. Some worry that principal-agent problems are particularly dangerous in a cyber bureaucracy [48]. Victims of extortion may fear that giving in will merely encourage the challenger to demand more, or that the threatened harm will be inflicted even after appeasement (e.g. the interrogator might kill the prisoner even after he confesses). Conversely, an actor cannot commit to carry out a threat if she would suffer more by administering the punishment than the disputed object is worth. Credible nuclear threats, for instance, are hard to make because the costs of Armageddon outweigh any imaginable political benefit. Commitment problems can sometimes be addressed through institutions (e.g. arms control regimes) that improve the monitoring and enforcement of agreements [49]. However, actors have trouble forming or sticking to an agreement if they have strong incentives to break it.

Uncertainty and commitment problems often appear together. Schelling [50] points out that uncertainty may restore some credibility to excessively costly punishments (i.e. "the threat that leaves something to chance" or rationally feigning madness like The Joker in *Batman*). Ironically, the possibility that an absolutely noncredible threat might not be exercised discounts its severity and enables an actor to create threatening situations where it might be exercised. For the same reasons, ambiguity about minor punishments undermines the effectiveness of deterrence by eroding commitment – a major problem in cybersecurity discussed further below.

The inability to distinguish offensive from defensive forces is another source of uncertainty about whether actors seek to preserve or revise the status quo [51], which then becomes an obstacle to committing to arms control agreements [52]. If they were distinguishable an actor might credibly advertise deterrence by denial without threatening other status quo actors, who could then agree to draw down offensive stockpiles. Indistinguishability is a major problem in the cyber domain because defensive intelligence collection to clarify capabilities and interests is hard to differentiate from offensive reconnaissance and network subversion. Compromise for intelligence today could be disruptive attack tomorrow. This ambiguity is particularly dangerous if offense is believed to have important advantages over defense, as is widely, if erroneously, believed about cyberspace [53]. The most vulnerable internet nations like the USA or China have strong incentives to engage in secret espionage for national security and economic reasons, so they cannot credibly commit to cyber norms that require them to give up an exploitation capacity that benefits them both [54], even as fielding cyber weapons may make miscalculation more likely in a crisis between them [55].

The limits of deception

The implications of bargaining theory for cyber conflict seem grim. We should expect uncertainty and commitment to be especially problematic for cybersecurity because deception is a permissive condition for network intrusion [16]. Social engineering persuades gullible users to participate in their own exploitation. Malware executes legitimate commands to produce behavior designers and users do not expect. Obfuscation of the method of, and responsibility for, intrusion complicates efforts to tailor defenses to the threat. Uncertainty about secret threats, reticence to reveal them for signaling, and the ability to cheat on normative agreements makes the use of low-cost deception almost inevitable for actors with developed cyber capabilities.

Yet reliance on deception has two important implications for bargaining that limit the political utility of offensive cyber operations. First, deception will be more useful for pursuing some positive benefit today (i.e. attempting to change the balance of power via theft, espionage, disruption, or counterattack) rather than coercively threatening harm tomorrow. Software vulnerabilities consist of private information that an attacker cannot reveal to make specific and thus credible threats, lest the defender close them via patching, reconfiguration, or countermeasure. This means that cyber weapons are ineffective deterrents by themselves, as is often noted. Some general deterrent effect might be gained through a reputation for skill in cyberspace, as Edward Snowden arguably helped to improve for the National Security Agency, but an immediate deterrent threat ("I will attack this network with this effect unless you refrain from X") is self-defeating. A reputation for skill might also bolster the perceived effectiveness of "network centric" military forces, but such deterrence relies on "cross domain" synergy rather than cyberspace alone [56].

Likewise, compellence is unlikely to work very well, either. Criminal ransomware embargoes a computer until the user pays an extortionate fee, but it must reveal a capability and usually makes only modest demands. More ambitious cyber coercion, such as North Korea's attempt to frighten Sony into suspending release of *The Interview*, often relies on threats that are vague or hard to take seriously, and in this case mobilized a serious investigation by the US government [57, 58]. Again, skill in the cyber domain might support military operations performed for some coercive purpose such as wearing down the will to fight through attrition, but this too relies on "cross-domain" synergy. Similarly, the attacker may use other diplomatic or military means to discourage retaliation once the benefits of an ambiguous attack have been realized [59].

The unsuitability of cyber weapons used by themselves for coercion removes an entire class of signaling moves – mobilization, demonstration, limited attacks, etc. – that states have traditionally made with military forces. Malicious hacking can still usefully support the

remaining set of distributional moves – warfighting, subversion, espionage, etc. – that attempt to alter the outcome of a conflict. Cyber operations might also support a more complex cross-domain strategy that harnesses the signaling potential of other, traditional, military means.

Second, offensive deception to change the balance of power becomes a liability when attempted against targets that are heterogeneous and/or well defended. There are many targets of low value connected to the internet, and many of them are poorly defended. Criminal malware can be downloaded from the web for low or no cost, and some attack platforms require very little knowledge to use. In contrast, targets of higher value tend to be more complex and better defended, and victims will be more willing to expend the effort on investigating the intrusion once detected. Moreover, the attacker is interested in exploiting a particular target rather than just anything in a class [60], which requires much more effort to reconnoiter the target, tailor malware and tactics to the target, and careful monitoring during the attack. If the target is in a disconnected network (behind an air gap) or has specialized control software or hardware peripherals, then the operational costs mount further.

Uncertainty facilitates exploitation, but uncertainty is a problem for attackers, too. Clauswitzian "fog of war" in the form of miscommunication and malfunction complicates maneuver in globally distributed computer networks as much as on the physical battlefield, driving up the costs of ambitious deception. The difficulty increases if the defender employs deception as well in the form of disinformation, honeypots, traps, counterintelligence, and other active defenses [61, 62]. Planning, intelligence, and control failures can result in the failure of the intrusion to achieve the desired effect or a compromise leading to defeat and, potentially, retaliation. These are organizational capacities that require considerable investment and experience even if the technology is cheap and widely available [63]. Furthermore, a successful exfiltration does not automatically result in a favorable change in the balance of power if the attacker is unable to convert intelligence into a competitive product or political outcome, or if the adversary compensates for the loss [64, 65]. Complexity begins to work against the attacker as target scale and heterogeneity complicates target reconnaissance and exploitation control, and as translation of the intrusion into a politically and economically meaningful result requires additional work and coordination. The increasing costs of attack against valuable targets offers some hope that strategies of denial can protect vital systems. The vulnerability of anonymous attackers to compromise in the most complex targets also offers some hope for deterrence strategies.

Deception exacerbates information symmetries and commitment problems, which are known to promote war. The prevalence of these problems in cyberspace explains the empirical epidemic of intrusions and the pessimism about deterrence. Fortunately, deception is useful for only a subset of political aggression, and it does not scale. The attribution problem – which turns on the ability to unmask deception – has been lurking throughout this discussion.

Attribution and deterrence

Uncertainty about capabilities and interest has been analyzed extensively in the IR literature, but the attribution problem seems to introduce a new form of uncertainty about the very identity of the opponent. Anonymity is not unique to the cyber domain, of course, as covert action, intelligence collection, anonymous extortion, guerilla warfare, and some offensive space operations all rely on the concealment of identity [66]. Muggers disguise themselves in order to demand money for protection from their own knife. Insurgents hide in the population in order to participate in rebellion demanding territory or regime change. What they have in common is a need to avoid near-certain punishment by a more powerful adversary – the police or the sovereign's army. Strong actors may also want to obfuscate their identity, for instance when an intelligence service conducts an assassination overseas, in order to avoid embarrassing questions or unwanted penalties from foreign allies or domestic constituents.

The attribution problem permeates the cybersecurity literature. An important recent study notes that "the attribution debate is evolving surprisingly slowly" with an excessive focus on technical forensics [15]. The authors argue that attribution is not impossible for the defender because even the most sophisticated attackers make mistakes eventually, but it is complicated, requires specialized skills, and is often time consuming. Moreover, the investigatory process can look to technical forensics as well as other intelligence sources and situational context [67], expanding the clues about means, motive, and opportunity that are available. Some confusion emerges from the fact that the identity to be attributed might refer to things as different as the location of the attacking infrastructure, the identity of individual operators or their organization, the originating and transiting jurisdictions, or the government that authorized or encouraged the attack [68], but all of these entities are also sources of clues that an intrusion might leave behind.

Anonymous bargaining

Plausible deniability is in fact so prevalent in security affairs that it is surprising it does not receive more attention in the conflict bargaining literature. Importantly, however, anonymity does not change the fact that there exists an opponent to bargain with. When we ask, "Who did this?" and, "How bad is it?" we at least know that someone has done something. The very abstractness of the bargaining model – player A vs. player B – becomes a virtue in this case. The strategic effect of anonymity is to inject additional uncertainty about the familiar factors that affect bargaining performance: the balance of capabilities, the demands and resolve of the opponent, and the costs and likely outcome of a conflict. Doing so both relies on and facilitates deception, which, as argued above, makes cyber conflict more likely. The attribution problem especially exacerbates bargaining failures for inexpensive, low-risk attacks, where the pool of potential miscreants is enormous. For more significant harms, however, fewer actors have the requisite capabilities and fewer still are resolved enough to realize benefit from major cyber attack. Anonymity, like deception in general, is more difficult to maintain at scale.

The perfect crime would conceal the injury as well as the culprit. Indeed, this is the goal of clandestine intelligence collection. The theft that goes undetected, the sabotage assumed to be an accident, or the mole everyone trusts can subtly change the distribution of power without the victim knowing that it has been changed. The resulting information asymmetry can adversely affect future bargaining performance, namely by making failure more likely as the victim overestimates his strength. Disinformation gambits, skillfully executed, might lead the victim to make poor decisions based on a false understanding of costs and benefits. For example, defenders may have false confidence in the ability of their air defenses to repel an enemy strike, ignorant of the fact that their radar systems have been compromised by code designed to disable early warning during enemy ingress. Of course, exercising this option, or any tactical surprise, requires the cyber attacker to reveal its prior coup. Purely covert exploitation depends on remaining undetected, changing the distribution of power but not conveying any true information. Purely anonymous coercion is almost impossible because communicating and understanding the power to hurt implies that there

someone doing the hurting and a target concerned about avoiding getting hurt.

Yet even the unwitting victim poses an implicit threat of compromise to the attacker, constraining his choices and strategy. Sensitivity to compromise is just another way of saying that the perpetrator is deterred from taking certain actions that would compromise the operation or expose the perpetrator to punishment. The term "self-deterrence" is sometimes used to describe self-imposed limits in the face of undesirable outcomes even if no one has issued a particular threat to impose retaliatory costs. Concern about collateral damage to civilian networks, fratricide against one's own systems, or the international opprobrium attached to the compromise of covert mischief may also encourage restraint [12]. As such, the sensitivity of the anonymous miscreant to compromise and unintended consequence has something in common with so-called general deterrence, which aims to prevent challenges of a given class and thus forestall crises, as contrasted with immediate deterrence, which threatens a specific target with consequences to prevent a specific action [69]. The antagonists cannot escape their bargaining relationship, even if one party is ignorant about the widening information asymmetry about the balance of power.

What is colloquially meant by the attribution problem arises when some intrusion or harm has been detected but the perpetrator has not yet been identified. The primary effect of uncertainty about identity at this phase is to create questions about the feasibility or desirability of retaliation. When the true abilities and interests of the attacker are unclear, the potential for future harm is uncertain; this creates uncertainty about the challenger's demands, which in turn undermines attempts to counter them by raising costs. A weak actor can feign strength, since the defender fears that mistaking a stronger target for a weaker one could lead to reprisals. Retaliating against the wrong target could be worse than failing to punish the right one, especially if it provokes a costly counteraction.

If plausible deniability and deceptive tactics make intrusions possible, then it follows from the bargaining model of war that they can be curtailed by clarifying the costs and benefits of continuing or ending the intrusion. Fighting can produce credible signals of resolve during war, just as militarized threats can inform during a crisis, but the act of fighting itself may even be more believable because combatants must absorb costs for an issue they care about. Limited wars provide information about the costs of a more total war [70], and fighting ends when the balance of power is clarified for both parties and they can commit to stop fighting [71]. Intrusion detection, active defense, and incident response are part of a negotiation "by other means" whereby the defender signals heightened costs or diminished benefits for the intruder. Solving, or at least attenuating, the attribution problem is an important part of this process. This interaction is extremely sensitive to not only technical architecture but also political and organizational context [15]. Attribution investigations chip away at the uncertainty attackers exploit by clarifying how and why systems were compromised. Defeating the deception upon which the attacker depends creates the possibility for tacit or explicit agreement to end the intrusion and to discourage future intrusions.

Note that if ambiguity about identity appears to undermine the credulity of deterrence, it also undermines compellence. The communication of a threat announces the presence of an attacker, whoever they are, which sets off the defensive search to identify the extent of the intrusion and raise defenses. That announcement may include real harm, as in wiping servers or doxing (releasing embarrassing documents), but then the damage is done and the ability to inflict more harm is reduced. It is still possible for the attacker to inflict harm after detection, say by threatening to reveal previously stolen embarrassing material from a cache, but it is not possible to obtain more documents through the same compromised vector, assuming the defender plugs the hole (I am grateful to Ben Buchanan for clarifying this point). Additional coercive demonstrations to improve the credibility of the threat provide more information for attribution, which opens up the attacker to reprisal, which he was presumably trying to avoid through anonymity. It is striking that the North Korean attacks on Sony ended right after the FBI announced it was investigating the breach, followed by an unprecedented statement by the President of the USA that attributed a cyber attack to a nation state and promised consequences. Reliance on anonymity is self-defeating in compellence because it signals vulnerability to retaliation.

Attribution vs. punishment

Attribution is not the same as punishment. It might be if the attacker really cares about a reputation for integrity: concerns about plausible deniability in covert action often turn on executive worries about blowback from domestic constituents, political rivals, and friendly allies as much as sanctioning by the victim. However, since retaliation can be costly for the one administering the punishment as well as the punished, a victim may decide not to do anything even after attribution if the costs and risks of punishing are too great. China conducts pervasive cyber espionage against Western interests, and there is good evidence available publicly and in US intelligence circles that the Chinese state is responsible, but so far naming and shaming actions have had little effect. The USA has had no stomach for more serious sanctions thus far, even as it routinely attributes Chinese (or Russian or Iranian) intrusions.

Furthermore, attribution is not required for punishment. A minefield, if advertised, protects an area from intruders whether they are known to the defender or not, since the trespasser selects a very specific punishment by the act of stepping on a mine. Defensive deception similarly turns the attacker's ability to penetrate a system against the attacker, by enabling the intruder's effort to trigger his own punishment or simply creating paranoia that he may have done so [16]. States often punish whole villages or classes of people when individual guerrillas cannot be brought to justice. This is not only unpopular but also expensive because a lot more force must be expended over a larger area. Indiscriminate punishment is usually a less effective deterrent or warfighting instrument in civil war than targeting killing [72], and may even counterproductively embolden insurgents, but it certainly does not depend on attribution. Likewise, a victim of cyber attack who wanted to punish some attackers could cease digital transactions with all trading partners or start attacking all suspects. Doing so, however, would quickly undermine the benefits that made internet exchange worthwhile in the first place and would be very unpopular. Inflicting even small punishments on a large number of suspects can be difficult to execute (so to speak) as well as politically illegitimate.

The belief that attribution is necessary for punishment thus turns more on the prohibitive political and social costs of indiscriminate retaliation. Uncertainty complicates the evidence collection process as investigators expend effort following false leads, allowing time to elapse and pressure for response to subside. An attribution case that depends on secret sources and methods, moreover, cannot be publicly revealed without jeopardizing those sources. An unconvincing attribution case, even if nominally correct, can undermine the legitimacy of a retaliatory act in the eyes of skeptics, especially in a democratic constituency, and especially if a punishment appears to violate norms of discrimination and proportionality. Negative audience costs might mitigate against doing anything at all to disturb the normal operation of the internet or provoke a reaction from the target.

In contrast, a more confident attribution case lowers these costs, because there are fewer targets to punish and punishing becomes more legitimate in front of political audiences. Narrowing the suspect pool simplifies the attribution investigation. Convincing public revelation of attributable harm can introduce further pressure to act on that information, provided policymakers care about the reputation costs of not acting. The public audience costs of inaction include emboldening future attackers as well as public anger at exploitation. These are more likely to kick in for more valuable targets where there will be more public pressure on the defender to respond to large insults, while the punishment will simultaneously be more concentrated on the perpetrator. Penalties for inaction or political rewards for action can thereby offset other aspects of defensive costs that are rising with target complexity. All other things being equal, attribution raises the probability of punishment, since most defenders will prefer legitimate and targeted punishments.

These considerations have nothing to do with cyber defense per se but with the cost of legitimating and enforcing the actual punishment by whatever means. Cross-domain punishments that are not as sensitive to revelation (e.g., nuclear missiles and conventional armies can be displayed in parades without undermining their potency) may further reinforce the credibility of threats to defend particularly highly valued targets. This "highly valued" caveat is important, as disproportionate responses will not be credible, and attackers, knowing this, will tend to tailor their attacks to fall below the threshold of response.

The sunk costs of attribution and the relative costs of retaliation

Attribution requires great technical expertise, analytical skill, and organizational coordination, and these require a lot of time, effort, and investment to develop. However, such costs will have already been paid in advance when a policymaker must decide to retaliate or not. The credibility of deterrence turns not on the substantial price tag of ex-ante attribution capacity but on the ex-post cost of administering punishment relative to the value of the resource protected. Attribution capacity and investigations may themselves be costly, but if successful they help to lower the potentially more significant political costs of administering punishment. The costs of developing the ability to attribute cyber attacks are sunk.

The states with the most valuable cyber targets are also the ones most likely to make the ex-ante investment in attribution capacity: "the larger a government's technical prowess, and the larger the pool of talent and skills at its disposal, the higher will be that state's ability to hide its own covert operations, uncover others, and respond accordingly" [15]. Advanced industrial states have the most lucrative networked economic and military targets, and they are the same ones that have the most sophisticated signals intelligence agencies to guard them. Moreover, only similarly well-endowed attackers will be able to pay the high costs of intrusion, which significantly reduces the suspect pool if the intrusion is detected. Finally, "the market for attribution has grown significantly" [15] thereby creating synergies between public and private attribution efforts, so this economy-wide investment can potentially create increasing returns especially for the most interesting targets.

Paying the high costs of attribution capacity in advance helps tremendously to lower the costs of actually carrying out the investigation after intrusion detection. Sinking costs generally can help to signal resolve [73]. Advertising the investment in attribution, likewise, can clarify for the attacker the declining costs of retaliation for the defender. The costs that are not sunk include incident response, forensic investigation, and the fusion of other sources of intelligence; this activity takes time and effort even if they fail to yield results. Yet such costs need not necessarily scale with target value and might even decrease, since the pool of suspects with the requisite means, motive, and opportunity will be smaller for more ambitious attacks. Moreover, there are likely to be more eyes looking for attackers once sophisticated intrusions are detected. The target might have to pay very little for assistance if allies, other government agencies, and private cybersecurity firms move in and start sharing information. The Iranian government paid nothing to the Western world for all the free open-source analysis of the Stuxnet worm and the investigation by journalists. If attribution and enforcement costs decline relative to the value of the systems being protected, threats to retaliate can be made more credibly.

The attribution problem is really hard where it less vital to solve it, but easier where it is most important. It follows that defenders will not invest equally in attribution, and attackers will not invest equally in hiding, for all types of intrusions. Investment will vary with the stakes of the bargain for each side. Intrusions that inflict greater harm will receive more investigatory attention and face higher likelihood of compromise and therefore punishment. At the same time, intrusions that inflict less harm are less likely to be attributed or punished. The feasibility of attribution is inversely correlated with the feasibility of deception and thus, as discussed above, the dominance of offense over defense. When there are many low value targets and it is cheap to exploit them, then there are many potential suspects. In contrast when there are few high value targets and it is expensive to exploit them, the pool contracts to just those with the means and expertise to pull off a sophisticated attack. The increasing cost curve for the attacker actually depresses the cost curve for the defender's attributional search. Counterintuitively, these tend to scale favorably for the defender with the complexity and value of the target, which are usually correlated (Complexity complicates intelligence and attack engineering whether or not the target is intentionally defended. Valuable targets will tend to receive more investment in attack prevention and mitigation whether or not they are complex. This correlation is not analytically necessary but is typical in practice).

The attribution problem exacerbates the bargaining failures that make cyber operations an attractive means for marginally altering the distribution of benefits in a political system through intelligence collection and covert disruption, but not for signaling resolve. Conversely, solving the attribution problem marginally reduces punishment costs and improves the ability to retaliate (although the risks of escalation and unintended consequences may remain after attribution). The well-understood results that flow from uncertainty and incredibility in IR theory generally are thus also relevant for a factor like plausible deniability that both increases ambiguity and undermines credible threats. Uncertainty about identity, not just capability and resolve, may appear theoretically novel, but its strategic relevance is expressed through its effect on familiar problems of information asymmetry and commitment. No additional conceptual machinery is needed in the basic bargaining framework to handle attributional uncertainty. Rather, attribution problems indirectly affect other things that affect deterrence. Once we understand these costs theoretically, then introducing attribution analytically becomes a matter of mapping concepts from a more complicated world to the simpler concepts we can model to improve understanding.

Modeling deterrence relative to target value

This section uses a formal model to show how different assumptions about costs and uncertainty affect the coverage and effectiveness of

deterrence by denial and punishment. I leave out a lot of complexity including signaling about costs sunk in cyber capabilities well in advance of any attack, the repeated interactions between attacker and defender during the course of the intrusion and attribution process, or decisions by the attacker to retaliate after being punished. The goal here is to show how arguments about the inevitability of deterrence failure in cyberspace can be challenged by examining simple assumptions about costs and uncertainty in deterrence more generally. I highlight some basic concepts, illustrate the dismal implications of the conventional wisdom about cyberspace, and finally show how different assumptions can restore the viability of deterrence. I work through the intuitions here and leave the details for the Appendix.

Denial and punishment

In this one-shot game the attacker S_1 can try to steal some benefit valued at v from the defender S_2 at the risk of getting punished for the attempt. The anonymous attacker makes no threat in advance (because cyberspace is ill-suited to coercion) while the defender publicly promises some retaliation r if attacked. There is some cost c_1 associated with the exploitation attempt, whether or not it is successful, and some cost c_2 to administer the punishment. These costs are initially fixed independently of v and r, but these assumptions will be relaxed later. If S_1 decides to exploit and S_2 does nothing, then S_1 wins v at the cost of c_1 and S_2 gets nothing. If S_2 retaliates, then S_1 has to pay r in addition to the c_1 he already paid while S_2 retains v but also has to pay c_2 to retaliate.

Note that c_2 only includes the costs of administering the punishments, not the full costs for defense. Administration costs may include the attribution investigation for the exploitation itself, convincing colleagues or the public of the legitimacy of punishment with a dubious attribution case, and the risk of escalation by the target if it turns out to be a powerful state. Attribution capacity of course depends largely on prior investment in expertise and organizational processes, and that prior investment may dwarf the costs of punishment administration in any given case. This is omitted in this model because defensive investments are sunk prior to this interaction and would be irrelevant to S_2's immediate calculation about whether or not to retaliate. The effectiveness of defensive investments, moreover, including attribution capacity, will be reflected indirectly in S_1's exploitation costs.

This simple model already captures deterrence by denial, the threat that attacker costs will be prohibitive, and deterrence by punishment, the threat that retaliation will offset the gains of exploitation. Figure 1 depicts the trough of deterrence failure between the two plateaus of protection by denial and punishment, respectively (For illustration in Figs 1 and 2, $c_1 = 0.2$ and $c_2 = 0.5$). Deterrence effectiveness (the vertical axis in Fig. 1) is modeled as the ratio of the protection premium – the costs of exploitation plus the risk of punishment – to the total value of the target. Perfect effectiveness means that no attacker chooses to exploit at that level because the target is not worth the high costs of exploitation or risk of retaliation. Anything less indicates some exposure of the target to attack. Large values imply protection from all but the most aggressive attackers. Small values imply exposure to less resolved types. With excessive operational costs ($c_1 \geq v$) exploitation is not profitable for S_1 so the lowest value targets are left alone. Assuming perfect information (for now), S_1 also knows that S_2 will be willing to punish if the costs of retaliation do not exceed the target's value ($v > c_2$). Deterrence thus fails only for targets valued in the range $c_1 < v \leq c_2$, the failure trough. Note that with perfect information the magnitude of the

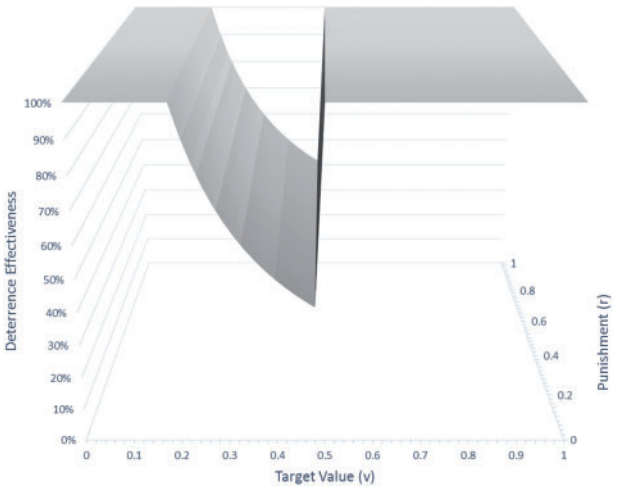

Figure 1. The trough of deterrence failure with perfect information.

punishment is irrelevant because S_2 keeps the resource with any decision to retaliate while S_1 still has to pay $r + c_1$ (I assume throughout that the punishment is not somehow a large reward, i.e. $r > -c_1$. As we shall see below the magnitude matters more with uncertainty).

The cliff on the left side of the trough of failure is the limit of perfect deterrence by denial. It begins to drop off as the ratio of target value to exploitation costs becomes more favorable for the attacker. The trough is deepest where the attacker's expected utility is maximized (The most profitable exploitations should not be confused with the most numerous. There are still costs of exploitation at this optimal level. If there is a large population of attackers who cannot afford them, they will be deterred by denial, and there will be more low value attacks). With perfect information, this point is right up against the defender's credibility threshold, the steep cliff on the right side which bounds deterrence by punishment. Given the costs S_2 must pay to retaliate relative to target value short of the credibility cliff, S_2 is more likely to prefer to forego punishment and cede the target. Under perfect information, S_2's bright red lines encourage S_1 to go right up to them (i.e. salami-slicing tactics). The clarity of this red line also makes the attacker insensitive to the magnitude of the punishment, since it is easily avoided. On the other side of the threshold, however, even a small punishment is worse than nothing since the defender keeps the full value of the prize, meaning that the highest value targets receive very good protection.

The trough is the essence of the so-called stability–instability paradox, where disincentives against major attacks incentivize lesser forms of aggression [74]. The bigger the trough and the steeper the credibility cliff, the more intense the paradox. The communication of credible information about thresholds and costs improves deterrence, yet it also incentivizes below-threshold aggression and other provocative risk taking.

Ambiguous threats

Policymakers often make ambiguous threats to improve deterrence in the trough, and also to leave their options open. Yet declining to commit – failing to signal resolve through "burning bridges"/"tying hands" or "burning money"/"sinking costs" [73] – creates a problem for credibility. Keeping options open signals that the defender might *not* carry through on a threat. Ambiguity is as liable to embolden the rash as it is to discourage the cautious.

As we relax the unrealistic assumption of perfect information, the magnitude of the threatened punishment begins to matter, and the trough opens up (Fig. 2). Risks of punishment for lower valued intrusions or the risks of higher punishments can dissuade exploitation, which expands the coverage of deterrence to some of the targets in the previous trough of Fig. 1. Yet if the attacker believes that the probability of retaliation p_r is low enough or the punishment is small enough, he may still be worth taking a chance, which erodes the effectiveness of deterrence for some of the higher valued targets covered by punishment in Fig. 1. S_1's estimate of p_r depends on what he thinks S_2's payoff to retaliation to be (v-c_2), and lots of errors (including defensive deceptions) can enter into his intelligence process; S_1 may also have some bias or risk acceptance that shifts his estimate to the right [Uncertainty is modeled as $p_r = \Phi_{\mu,\sigma}(v - c_2)$, as discussed in the appendix. For illustration in Fig. 2 the estimative error $\sigma = 0.1$ and the bias $\mu = 0.1$]. The new ambiguity of the threat erodes the credibility cliff and enables more aggressive attackers to exploit higher value targets, especially when the consequences threatened are insufficiently painful. The stability–instability paradox persists, but with an expanding gray zone where credible and sufficient retaliation is in doubt. The defender must, therefore, decide if she wants to defend higher value targets more effectively by making clear threats, or cover more targets less effectively by making ambiguous ones.

Note that misperception by the attacker can have the same effect. There are two different ways to interpret the increased uncertainty about the probability of retaliation. First, uncertainty might be the result of an intentionally ambiguous threat made by the defender, or a bluff that is suspected by the attacker. This has the effect of broadening deterrence but weakening it where it counts most. In that case Fig. 2 is a picture of reality, and the attacker can move through the trough and up to its edge without being punished. Second, however, the uncertainty might result from misperception by the attacker (or miscommunication by the defender) about the true balance of costs. Risk estimation is subjective from the attacker's perspective, and might be overly optimistic. In that case, Fig. 2 is the attacker's perception and Fig. 1 is a picture of reality, which means that the attacker might exploit targets to the right of the defender's credibility cliff, drawing retaliation. This reflects the classic notion of bargaining failure via information asymmetry. Deterrence failure in either case entails a decision to exploit, regardless of whether or not the defender will punish.

So far this model has said little in particular about attribution, or even cyberspace. We can capture some of the conventional wisdom about the effects of cyberspace and its cheap anonymity on conflict by adjusting some of the key parameters. First, it is widely believed that cyberspace lowers the costs of attack by providing connections across borders, downloadable attack software, and plenty of opportunities to adopt aliases and surreptitious attack routing. Second, most people believe that is also raises the costs of network defense because of coordination problems, infrastructural inertia and network effects, and a growing attack surface. Third, punishments for cyber crimes often seem insufficient because even if an intrusion is attributed, the knowledge that the perpetrator is located in a foreign jurisdiction or is a foreign state makes large punishments unattractive or infeasible. Fourth, the many layers and players in cyberspace creates much uncertainty for the assessment of costs. Fifth, the cheap proliferation of cyber tools means that there will certainly be some risk accepting attackers in the pool. Figure 3 shows that these assumptions create a disaster for deterrence (For illustration in Fig. 3, $c_1 = 0.05$, $c_2 = 0.6$. For all of the cyber deterrence examples – Figs 3–6 – uncertainty is set slightly higher with some risk acceptance in the attacking population, $\sigma = 0.15$ and $\mu = 0.15$).

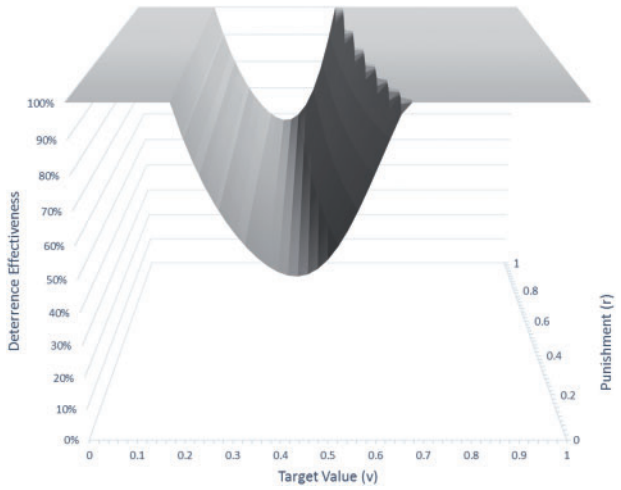

Figure 2. Uncertainty both improves and erodes deterrence.

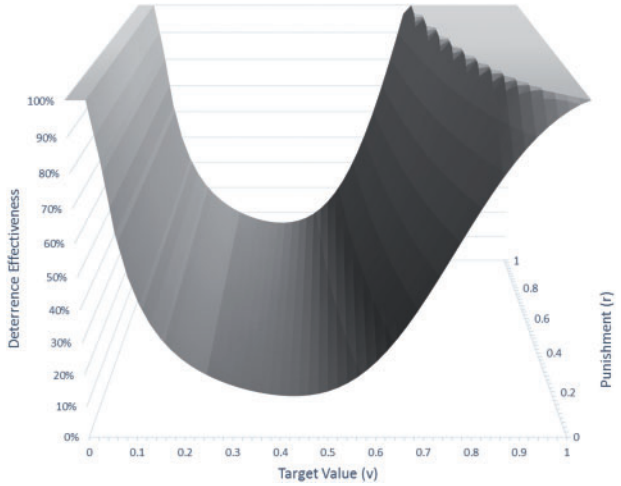

Figure 3. Low attacker costs, high defender costs, high uncertainty.

The trough of failure is deep and wide. Reduced attacker costs bite into deterrence by denial and shrink the plateau on the left, also creating a steeper cliff as defenses fail moving right. Higher defender costs and uncertainty about them, and lower punishments and great willingness to risk them, bite into deterrence by punishment on the right. Most of the value of the target set is exposed to exploitation.

This unhappy outcome is not far from what many cyber pessimists expect: the erosion of deterrence exposes more and more of the most valuable systems to attack from newly empowered hackers. The only hope for marginal improvement offered by Fig. 3 is to increase the punishment, which restores the plateau protection for the most valuable targets on the right. With insufficient punishments, the costs of retaliation are too high and there is no credible deterrent for the entire set of targets. It is therefore unsurprising that many policy statements about deterring major cyber attacks stress the possibility of a military response not limited to the cyber domain [75]. Cross domain capabilities that are less sensitive to revelation than cyber weapons alone enable S_2 to announce the possibility of a large punishment in advance, which raises the stakes for the attacker and accounts for the gentle upward slope on the right side of the trough. Is it possible to improve further on this dismal situation?

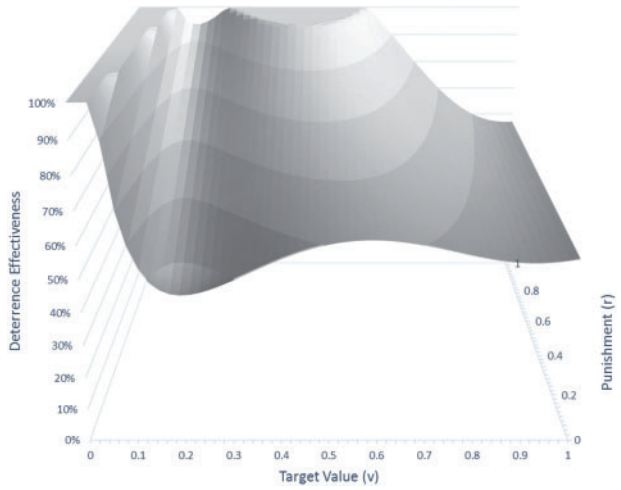

Figure 4. Defender costs scale faster than attacker costs.

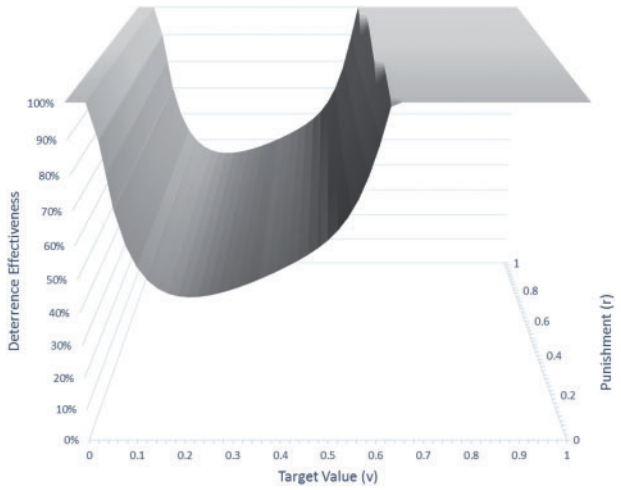

Figure 5. Defense costs falling as attacker costs rise.

Scaling costs

So far I have assumed that costs are fixed with respect to target value. As discussed in the previous section, this is totally unrealistic. The question is how do costs scale and what is the relative relationship between attacker costs of exploitation and defender costs of administering the punishment? It is not very controversial to assume that attacker costs will scale with target value. Target complexity and investment in defensive measures improves deterrence by denial by raising the costs of exploitation against more important targets. Defender costs pose a more complicated question. Many assume that cyberspace is categorically offense dominant, which means that the defender always has to spend more to defend the same target than the attacker has to spend to attack it. Thus defender costs might scale at least as fast as and probably faster than attacker costs. More complex targets would require more effort to defend because they have a larger attack surface and coordination problems among defenders are more severe.

Figure 4 suggests that the result is a more complicated disaster for deterrence, but still a disaster (Costs are modeled as $c_1 = c_{1\min} + \beta_1 v^{k_1}$; $c_2 = c_{2\min} + \beta_2 v^{k_2}$. For illustration in Fig. 4, $c_{1\min} = 0.05$, $\beta_1 = 0.5$, $k_1 = 2$, $c_{2\min} = 0.2$, $\beta_2 = 1$, $k_2 = 2$). The lower initial costs for S_2 shift her credibility cliff to the left until her exponentially increasing costs eventually exceed the linear increase in the value of the target, which in effect creates another trough on the right as soaring costs erode credible retaliation. The bump between these two troughs reflects the modest contribution of deterrence by punishment, which is sensitive to the magnitude of the punishment. It is deterrence by denial – the difference between target value and attack cost – that provides most of what protection does exist. Deterrence by denial makes the failure troughs in Fig. 4 shallower (better coverage) than in Fig. 3. Further, the right trough in Fig. 4 is shallower than the left, which means there is slightly better protection by denial for higher value targets, but they are still dangerously exposed (less effective). Figure 4 dramatically captures the pessimism many have regarding cyber deterrence, and, if one is forced to choose between two bad alternatives, the widely shared preference for denial strategies over punishment.

The fact that we have not seen many or any of the worst attacks might just mean we are drawing from one side of the distribution and we are due for a catastrophe. Or the model might be wrong. Fortunately, if we change the assumption about defender cost scaling to better align with the discussion of deception and attribution above, the situation does not appear so dire, nor does the history of moderation in international cybersecurity appear so improbable. One obvious objection to such scaling is that attribution remains a complex, resource-intensive, and time-consuming process, even or especially for higher value targets. Recall, however, that in this model c_2 is the cost S_2 has to pay to retaliate while most attribution capacity investments are paid prior to the game and thus do not figure in to the decision to retaliate or not. Capacity investment, which is paid whether or not S_2 retaliates for a given attack, can be expected to influence the shape of c_1 and c_2. It is these sunk costs that determine the size of the residual costs for punishment, and substantial investment to assure attribution for the highest value targets could greatly reduce the costs of investigation and legitimation of the punishment after the actual attack. A more complicated game might model this relationship explicitly.

Attribution becomes more feasible as the number of suspects with the ability and motivation to attack higher valued targets decreases. Furthermore, the complexity of higher valued targets creates a less permissive environment for attackers and increases the chance that they will make a mistake and leave clues for attribution. Contrary to conventional belief, the offense–defense balance changes at scale with some inflection point where the balance shifts from offensive to defensive advantage. Near where that shift occurs, and perhaps well before, deterrence becomes feasible again. Not only does attribution become more likely at scale, but also a convincing attribution case will lower the costs of delivering and legitimating retaliation.

In Fig. 5, the costs for the attacker start out very low and increase steeply with target value, while the costs for the defender start very high and falling off as attribution and punishment becomes less costly (For illustration in Figs 5 and 6, $c_{1\min} = 0.05$, $\beta_1 = 1$, $k_1 = 2$, $c_{2\min} = 0.9$, $\beta_2 = -0.9$, $k_2 = 2$. Uncertainty is the same as Figs 3 and 4). Thus offense dominates for lower valued targets but defense dominates for higher ones. The result for deterrence by punishment is more encouraging, with the right side plateau covering high value targets. Furthermore, as the exponentially increasing attacker cost curve exceeds the value of the target, deterrence by denial again becomes a factor after having dropped off the left side plateau. The contribution of deterrence by denial is reflected in the gradually increasing slope of the floor of the trough before the credibility cliff. Deterrence becomes more feasible with a dwindling offensive advantage at scale. The stability–instability effect is not eliminated, as the failure trough still exists, but it is shifted left and up considerably compared to Fig. 3.

The model in Fig. 5 shows an insensitivity to the magnitude of the punishment, which seems unrealistic. The value of the target, not the magnitude of the punishment, is the controlling factor in the credibility of the threat in this particular model. Injecting additional uncertainty into the attacker's estimate (not shown) would restore some punishment-sensitivity, but not much. To introduce more realism we can further make retaliatory costs proportional to the magnitude of the punishment. The larger the retaliation, the great the risk of escalation and potentially the cost of administering punishment in the form of military response. Figure 6 introduces this additional scaling factor so that retaliatory costs scale inversely with target value, as in Fig. 5, but proportional to punishment (Defender costs can be modeled $c_2 = c_{2\min} + \beta_2 v^{k_2} + \beta_3 r^{k_3}$. For illustration in Fig. 6, $\beta_3 = 0.9$, $k_3 = 2$, all other values the same as Fig. 5). The result is that the higher magnitude of threatened punishment actually bites into the credibility threshold, eroding the punishment plateau. Lower punishments, ironically, provide more protection, assuming per model specifications that the attacker is denied the benefit of the target any punishment leaves him worse off. The most extreme punishments are credible only for the most valuable targets.

Figure 6 shows how different scaling assumptions are possible, with different consequences for coverage and effectiveness of deterrence. One should be very cautious when interpreting these results, bearing in mind that the model assumes no value transfer to the attacker if the defender retaliates, and no formal signaling moves via sunk cost investment or other means before the decision to attack. In the case of intellectual property theft or intelligence collection, the detected exploitation might still transfer the value, thus the attacker might be willing to suffer the punishment. In that case a larger punishment would be needed to compensate. The point of the analysis here is merely to show how scaling assumptions can matter – determining the actual shape of the scaling function is a much more complicated analysis depending on the nature of the actual targets to be protected from attacks of a certain character with policy tools having particular costs. Such analysis is beyond the scope of this article.

Figure 5 (as well as Fig. 6 with the caveats noted above) is a much better match to the empirical distribution of cyber attacks we experience in the real world than any of the previous figures. The majority of computer users do not actually experience that much cyber crime because their data are not really worth that much. For slightly more lucrative but still low value targets, there is an epidemic of cyber crime. More ambitious intrusions target more valuable targets but at a diminishing rate. At the high end, there is no activity because not even the actors with the ability to launch a digital Pearl Harbor have the motivation to do so. Deterrence works, just not everywhere. We see a lot of little attacks, and little of a lot. When assumptions about cost scaling are adjusted consistent with the discussion of deception, attribution, and punishment above, we can be slightly more optimistic about the viability of deterrence for what matters.

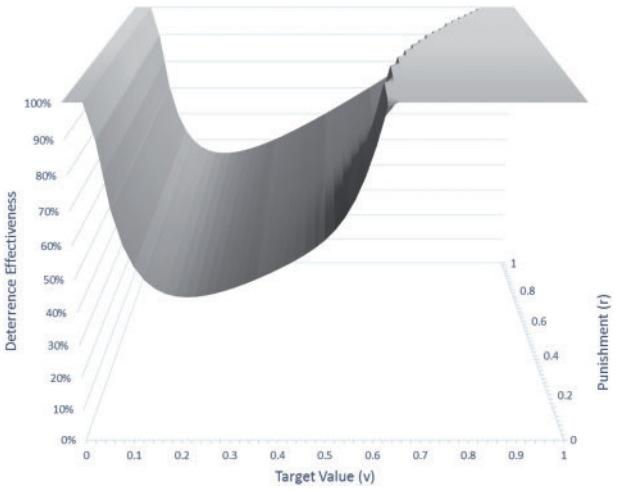

Figure 6. Retaliatory costs scale inversely with target value and proportional to punishment magnitude.

Conclusion

Technology does not create political imperatives by itself [76–78], and cyberspace is no exception. Arguments about strategy in cyberspace tend to be posed deterministically, as if particular technical features of computer networks make certain strategic outcomes inevitable. Does offense dominate defense? Is attribution unreliable? Is deterrence doomed? No absolute answers follow from the nature of computing itself. Deterrence strategies can be effective for protecting high value targets where defender resolve is high and attribution is likely. For targets where retaliatory threats are less credible and attribution is more difficult, however, denial and defense offer better protection. Defensive deception, furthermore, can improve both deterrence and denial, but it is more likely to be used to protect more valuable targets, reinforcing the scale-dependent effectiveness and coverage of deterrence.

The means of conflict in cyberspace seem revolutionary but the nature of conflict there looks familiar from an IR perspective. Aggression can range from total war for existential survival to capricious bullying. Clausewitz notes that "wars can have all degrees of importance and intensity, ranging from a war of extermination down to simple armed observation" [33]. Moreover, the most extreme "conflict will not occur very often, for if the motivations are so powerful there must be a policy of proportionate magnitude. On the other hand, if policy is directed only toward minor objectives, the emotions of the masses will be little stirred" [33]. So too in cyberspace, service abuse and data theft abound but serious disruption is rare. Some of the most notable events to date, such as the Russian distributed denial of service (DDoS) attacks against Estonia in 2007 and Georgia in 2008, the Stuxnet attack on Iranian nuclear infrastructure discovered in 2010, Iranian reprisals against Saudi Aramco in 2012, or the North Korean extortion of Sony in late 2014, are notable for their lack of violence and surprising restraint. Estonia did not replace its Soviet statue, Russia did not need cyber attacks to defeat Georgia, Iran increased its output of enriched uranium during and after Stuxnet, Aramco replaced its computers and kept on pumping oil, and millions of Americans watched *The Interview*. Deterrence failures in cyberspace are notably skewed toward targets of minor value where the damage is temporary and remediable, and where the adversary has additional work to do beyond a successful intrusion to realize any real shift in the distribution of power. Meanwhile, the emotions of the masses have been little stirred by the constant irritation of cybercrime and espionage, much to the frustration of advocates for more robust cyber policy.

The 2011 Department of Defense Strategy for Operating in Cyberspace put little stock in deterrence. It focused mostly on defense and denial because of a lack of confidence about attributive capacity. Four years later the situation seems different. The 2015 DoD Cyber Strategy advertises its improvements in capacity and their application for deterrence (pp. 11–12):

> Attribution is a fundamental part of an effective cyber deterrence strategy as anonymity enables malicious cyber activity by state and

> non-state groups. On matters of intelligence, attribution, and warning, DoD and the intelligence community have invested significantly in all source collection, analysis, and dissemination capabilities, all of which reduce the anonymity of state and non-state actor activity in cyberspace. Intelligence and attribution capabilities help to unmask an actor's cyber persona, identify the attack's point of origin, and determine tactics, techniques, and procedures. Attribution enables the Defense Department or other agencies to conduct response and denial operations against an incoming cyberattack.

Advertising improvements in the ability to attribute, especially the availability of clues and context beyond technical forensics that simplify analysis, helps to improve deterrence. By the same token, disguising attributive capacity or failing to make a convincing case in public undermines both the credibility of deterrence and the legitimacy of the response. The skepticism of the commercial cybersecurity community in the wake of US government attribution of North Korean for hacking Sony, for example, might have been addressed more proactively by talking about some of the evidence available and the process of analysis in ways that avoided disclosing sensitive sources and methods.

Although the attribution problem becomes simpler at scale, solving it is insufficient for deterrence. The scale-sensitive nature of the attribution problem inverts the offense–defense balance for higher value targets in favor of the defender, but the problems of uncertainty about costs and risk remain in the gray zone of the credibility trough. This, unfortunately, is precisely where challengers focus their attention. For years major breeches have been attributed to China and US policymakers have threatened unspecified consequences but have been unwilling to take costly action [79]. Whatever the losses inflicted by cyber attackers – and the real impact of intelligence collection on the balance of power goes beyond the technical penetration and depends further on the ability to disseminate and apply whatever is collected – the costs of fighting back might be more. For instance, US sanctions for Chinese cyber attacks that result in the retaliatory exclusion of US firms from the Chinese market could be more damaging to US economic performance than the loss of intellectual property. President Obama has suggested the universal adoption of norms of nonuse, stating "there has to be a framework that is analogous to what we've done with nuclear power because nobody stands to gain" [80]. Yet the fact is that the USA has plenty to gain from computer network exploitation for national security intelligence in peacetime and tactical electronic warfare in wartime and would not be willing to restrict this capability – enforcing norms of nonuse would be too costly for the USA.

In recent testimony before Congress the US Director of National Intelligence stated, "The muted response by most victims to cyber attacks has created a permissive environment in which low-level attacks can be used as a coercive tool short of war, with relatively low risk of retaliation" [81]. He further opined, "Until such time as we do create both the substance and the mindset of deterrence, this sort of thing is going to continue … We will continue to see this until we create both the substance and the psychology of deterrence" [82]. This is easier said than done, however, because it would require the USA to pay some cost to deter. Note that Clapper does not say that deterrence is categorically impossible, just that most attacks fall below the threshold where it is credible. This is no accident but a strategic choice by attackers sensitive to that threshold. Shifting the threshold to cover less valuable targets would require the defender to somehow lower its costs of retaliation or the attacker's estimate of those costs. The alternative is to raise the uncertainty about whether a costly punishment will be enforced, in effect discounting the cost that must be paid, but this move carries the risk of exposing more valuable targets.

The unfortunate implication is that if policymakers really want effective deterrence against truly important targets, then they have to set priorities and be willing to tolerate exploitation in the gray zone against low and even medium value targets. Improved attribution and credible declaratory policy that specifies the types of targets to be protected, particularly by "cross domain" consequences that are more useful for signaling (i.e. robust to revelation), can be expected to improve protection from attacks that might cause major loss of life and damage to property. Yet clarity encourages the development and provocative use of lesser irritants, and also creates difficult tradeoffs across other policy priorities such as industrial efficiency and civil liberties. Ambiguous policies about whether a particular type or magnitude of punishment might be triggered by crossing some poorly delineated threshold are a double edged sword. Uncertainty increases the range of deterrence while degrading its quality.

A further unfortunate implication is that clarity for the punishment regime may require the mobilization of larger responses than are strictly necessary in order to deter the risk accepting attacker, i.e. "overkill" [84]. Attackers know that ambiguous threats are often used to cover targets that the defender cannot credibly punish. A little knowledge of past deterrence failures or risk acceptance on the part of attackers is enough to make them willing to call the bluff. If the risk accepting attacker decides to exploit a target beyond the actual threshold, believing a real but ambiguously communicated threat is a bluff, it can trigger a punishment that will be costly for the defender to administer. If attacks on a certain class of targets are absolutely unacceptable, the threatened consequences might have to be disproportionate to be believed. The signaling utility of overkill is yet another reason why threatening cross-domain retaliation for unacceptable cyber-attacks is sensible. It becomes ever more important to be clear about what "unacceptable" means. The analysis above suggests, however, that a major downside of threatening disproportionate response is that deterrence effectiveness for all but the most important targets is compromised.

Creating a regime for effective deterrence means accepting another regime where denial and defense will be necessary. Deterrence can be effective where it matters as long as policymakers accept that contestation elsewhere will be interminable. There will always be a gray zone where important targets – but not the most important – will be attacked by increasingly sophisticated adversaries. Defense in depth and counterintelligence deception strategies become important where attacks cannot be deterred. Moreover this activity is democratizing as firms and nongovernmental organizations increasingly contend with intrusions from state and non–state sources. Cyber defense becomes more complex and contests with ambiguous adversaries become more interminable as a result. The efficacy of deterrence in one area can never be an excuse for inaction in another, and network defenders will experience their job becoming more and not less difficult. Conversely, the evolving epidemic of cyber insecurity does not mean that cyber deterrence is infeasible in any absolute sense; on the contrary, threats are becoming more sophisticated because deterrence is working.

Acknowledgements

The author thanks Ben Buchanan, Shannon Carcelli, Allan Friedman, Erik Gartzke, Stephan Haggard, and the anonymous reviewers for valuable comments on earlier drafts. This research was supported by the Department of Defense Minerva Initiative and Office of Naval Research Grant [N00014-14-1-0071].

Appendix

Figure 7 depicts a one-shot deterrence game with some uncertainty in S_1's estimation of S_2's costs and target valuation. Parameters v, c_1, c_2

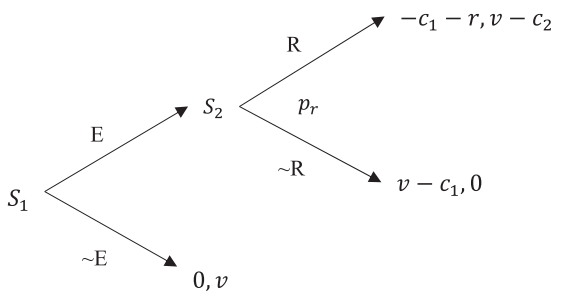

Figure 7. Simple deterrence game.

and r range on the real numbers [0,1]. S_2 gets to pick the value of r prior to the game, which S_1 knows in advance, and the values of c_1 and c_2 depend on sunk cost investments and other exogenous factors.

The target valuation is assumed to be symmetrical for both players. It is possible in the real world for the players to have different valuations of the target, but this greatly complicates the analysis without significantly changing the results of this single shot game without signaling. This potential difference can also be considered an additional source of uncertainty incorporated into the measure of the probability of retaliation with imperfect information, introduced below. For analytical convenience the defender keeps the entire target value on a decision to retaliate. It would be possible to include a lottery term whereby the attacker might get to keep something, like some intelligence collected, even if punished (although the intel might thereby become invalid!), but this does not seem to interestingly change the results.

The game is trivial with perfect information. If $c_1 \geq v$ then S_1 will not exploit, and S_2 retaliates if $v > c_2$. S_1 understands this and so exploits only in the range $c_1 < v \leq c_2$. However, if S_1 is unsure of v or c_2, S_1 will have to estimate the defender's payoff to retaliation. Assuming that estimative errors are normally distributed, then S_1's subjective estimate of p_r can reasonably be modeled with the normal cumulative distribution function on v-c_2, which is the threshold of credible retaliation.

$$p_r = \Phi_{\mu,\sigma}(v - c_2).$$

This is a slightly unusual way of modeling uncertainty. Game theorists often represent uncertainty by allowing "nature" to make a prior move with some probability, or with some parameter where only the distribution of values is known. Yet for a one shot game representing an attacker with a difficult intelligence and planning problem, it is reasonable to let the attacker try to make the best estimate possible, with some error rate exogenous to the interaction due to collection problems, bureaucratic errors, processing delays, defender deception, and other friction. These random errors are assumed for the purposes of modeling to be normally distributed. The standard distribution of errors becomes a convenient measure of how much noise there is in the attacker's estimative process. Further, shifting the mean on the error distribution is a geometrically convenient way to represent a systematic bias in over(under)estimating the credibility threshold, whether that is a result of risk acceptance (aversion) or something else. This is again a nonstandard way of modeling risk aversion, which is usually considered in terms of the utility curve or indifference between winning a lottery with some probability and having the same fraction of a good for sure. Here, however, we are interested in how the attacker's estimate of the probability of retaliation shapes his expected utility and systematically biases the estimate of defender credibility.

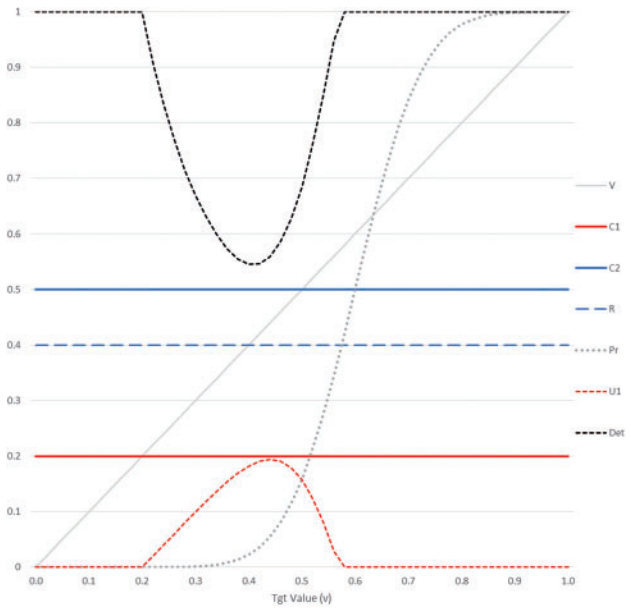

Figure 8. Deterrence with uncertainty.

For a well-informed attacker, the standard deviation σ will be very small and p_r looks like a step function at $v = c_2$ in the plane of these two parameters. If the attacker is more uncertain about these values, then σ is larger and p_r will be a more spread out S-curve, again inflecting at the actual $v = c_2$. Furthermore, the mean μ can be interpreted as a measure of S_1's estimative bias toward risk acceptance because it shifts the inflection point where S_1 guesses $v = c_2$. Risk-accepting attackers ($\mu > 0$) will tend to discount the defender's resolve by either overestimating c_2 or underestimating S_2's valuation of the target, shifting the mean of the distribution of errors, and thus the inflection point of the error curve, to the right. Risk averse attackers ($\mu < 0$) overestimate resolve and shift the curve left. The values of σ and μ are chosen exogenously in this one-shot game; however, in the real world players certainly might be able to adjust them ex-ante just as they adjust their sunk cost investments. For example, if S_1 tends to make ambiguous threats, then σ will surely be larger for S_2.

S_1 will thus attack only if expected utility $U_1 > 0$ where:

$$U_1 = p_r(-r - c_1) + (1 - p_r)(v - c_1) = v - (c_1 + p_r(v + r)).$$

U_1 can be thought of as the prize value less some premium$c_1 + p_r(v + r)$, which can be distinguished into a denial cost component and a punishment risk component. Because S_1 would be indifferent between taking only a discounted fraction of the prize at no cost and paying the premium to win the full prize, the premium can be thought of as the total level of protection for the target. The effectiveness of deterrence will here be defined as the ratio of the protection premium to the total value of the target. Perfect effectiveness ($Det = 1$) means that the attacker chooses not to exploit at that level. Anything less than 1 indicates some exposure of the target to attack.

$$\text{Det} = \begin{cases} \text{if } v > 0 \text{ and } U_1 > 0, & \dfrac{c_1 + p_r(v + r)}{v} \\ \text{otherwise,} & 1 \end{cases}.$$

Figure 8 illustrates how these various parameters with the same values as Fig. 2, still assuming costs are invariant with value. Note that the trough of deterrence failure is in the range $c_1 < v \leq c_2$, which can be visualized as the intersection of the cost curves along the diagonal

line in the figures, but is shifted to the right because of uncertainty and risk acceptance ($\sigma = 0.1$, $\mu = 0.1$) in the s-curve of p_r.

To allow these costs vary with target value, as discussed in the text, we redefine them as:

$$c_1 = c_{1min} + \beta_1 v^{k_1}$$
$$c_2 = c_{2min} + \beta_2 v^{k_2} + \beta_3 r^{k_3}.$$

The text provides numerical examples and discusses the implications of different scaling parameters β_n and k_n. Figure 9 is a companion to Fig. 4 that illustrates the interaction of these different terms. While the lower initial costs for S_2 move the credibility cliff to the left (i.e. where $c_2 = v$), the concave cost curve eventually crosses the diagonal value curve again, which in effect creates another trough.

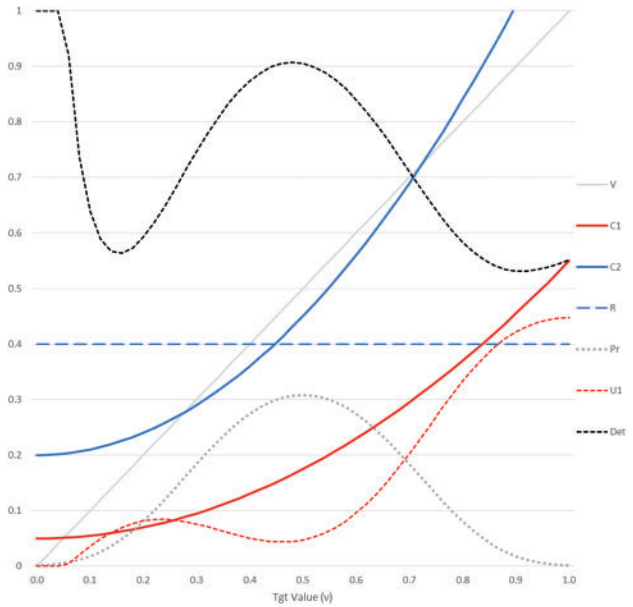

Figure 9. Scaling costs.

Note that the estimate for p_r is not even an S-curve anymore, for credibility is eroded as defender costs climb. If different parameter values are chosen so that the cost curves cross as in Fig. 5 the credibility cliff will be located at the attacker's (risk accepting) assessment p_r of where $v = c_2$, not necessarily where the offense-defense balance crosses over (at $c_1 = c_2$).

The simple game does not include specific moves for attribution. The assumptions about attribution in the text are captured indirectly through their effect on uncertainty and costs. Simple is often better if the goal is to understand underlying mechanisms rather than make predictions, and simple is usually more tractable. When the model does not behave like reality, then it is easier to find out why when there are fewer assumptions to juggle. Yet there is nothing sacrosanct about the form of a game. We could include a specific move for attribution, which may provide a more intuitive, if more analytically complicated, representation of the attribution problem.

Figure 10 is an extension of Fig. 7 with an additional step between the attacker's decision to exploit and the defender's decision to retaliate. This is a move by "nature" where the intrusion is detected and attributed with some probability p_a. Note that S_2 can retaliate even if she does not attribute the identity of S_1, but if so she pays an additional penalty (e.g. the costs of rounding up the usual suspects or the unpopularity of indiscriminate or categorical punishment); for convenience this is the same as the punishment r (All that is necessary is that the penalty be nonzero. Figure 10 illustrates and I discuss below that the $p_{r|\sim a}$ curve is simply shifted to the right of $p_{r|a}$. A more complicated but less intuitive specification of these costs could modify the shape of the credibility cliff somewhat, e.g., creating a cliff within a cliff, but this adds little to our understanding). As before, defensive investments are sunk prior to the game and not included explicitly in the model, although they are assumed affect the shape of the offensive and defensive cost curves.

The probability of attribution can be modeled as the balance of offensive and defensive investments (i.e. attribution is hard when defense has to pay relatively more than offense). As before, the attacker's estimate of the defender's conditional probabilities of retaliation can reasonably be modeled with the normal cumulative distribution function, with the standard deviation representing

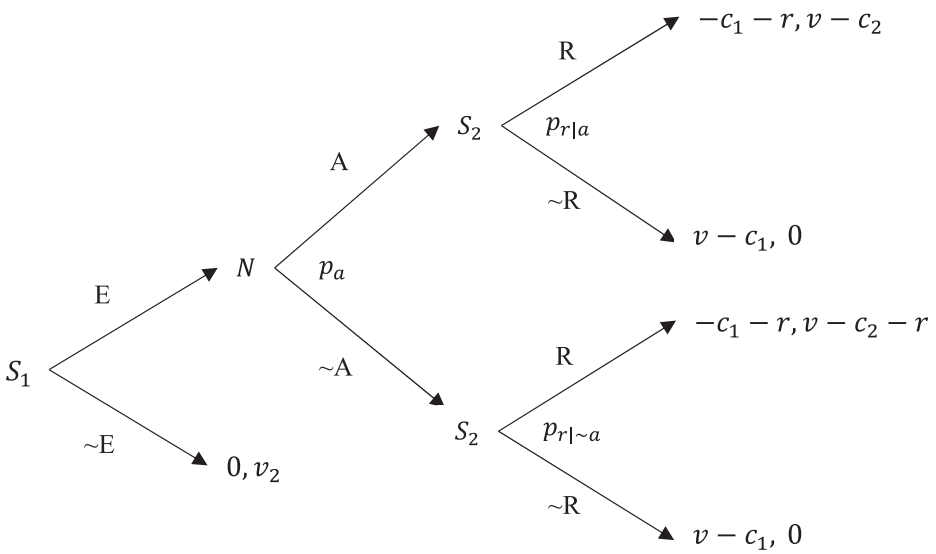

Figure 10. Modeling attribution explicitly.

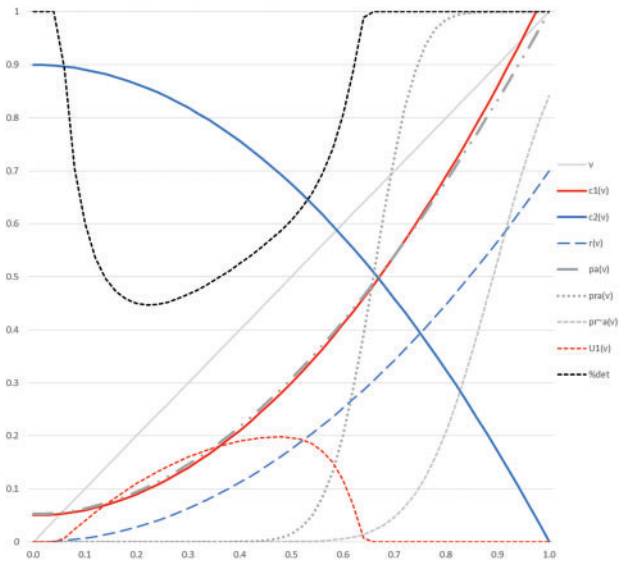

Figure 11. Deterrence with alternate specification.

uncertainty and the mean representing risk acceptance. The attacker's utility works out similarly as before, with a denial and a punishment discount, but the probability of punishment is a bit more complicated because of the defender's option of costly retaliation without attribution:

$$p_a = \frac{c_1}{c_1 + c_2}$$
$$p_{r|a} = \Phi_{\mu,\sigma}(v - c_2)$$
$$p_{r|\sim a} = \Phi_{\mu,\sigma}(v - c_2 - r)$$
$$U_1 = v - \left(c_1 + (p_{r|a}p_a + p_{r|\sim a}(1 - p_a))(v + r)\right).$$

Similar assumptions about c_1 and c_2 produce similar results. Figure 11 plots a numerical example with the same values as Fig. 5 with the additional probability curves with the assumptions that costs scale with value, inversely for attacker and defender (they scale in this example at different exponential rates to make the combined attribution probability visible on the graph, otherwise $p_a = c_1$), with uncertainty and risk accepting attackers. The result for deterrence is not substantively different from Fig. 5.

Indeed, varying the assumptions about costs appears to produce similar results across the two model specifications. In essence, the additional cost on the nonattributed punishment produces a higher credibility threshold than attributed punishment, and so it is largely ignored by the attacker under most conditions. This can be seen in the relative position of the s-shaped curves in Fig. 11. The additional modeling complexity of including an explicit step for attribution does not meaningfully change the results of the analysis, as long as we make similar assumptions about the relationship between attribution, costs, and uncertainty. For understanding the basic dynamics of bargaining under uncertainty, attribution is best understood as an exacerbating factor for more fundamental cost and interest uncertainties, as well as commitment problems.

References

1. Libicki MC. *Cyberdeterrence and Cyberwar*. Santa Monica, CA: RAND, 2009.
2. National Research Council (ed). *Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for U.S. Policy*. Washington, DC: National Academies Press, 2010.
3. Lupovici A. The 'Attribution Problem' and the social construction of 'violence': taking cyber deterrence literature a step forward. *Int Stud Perspect* 2014.
4. Snyder GH. *Deterrence and Defense: Toward a Theory of National Security*. Princeton, NJ: Princeton University Press, 1961.
5. Peterson D. Offensive cyber weapons: construction, development, and employment. *J Strat Stud* 2013;**36**: 120–4.
6. Solomon J. Cyberdeterrence between nation-states plausible strategy or a pipe dream? *Strat Stud Q* 2011;**5**:1–25.
7. Iasiello E. Is cyber deterrence an illusory course of action? *J Strat Secur* 2013;**7**:54–67.
8. Gartzke E. The myth of cyberwar: bringing war in cyberspace back down to earth. *Int Secur* 2013;**38**:41–73.
9. Elliott D. Deterring strategic cyberattack. *IEEE Secur & Privacy* 2011;**9**:36–40.
10. Denning DE. Rethinking the cyber domain and deterrence. *Joint Forces Q* 2015;**1**:8–15.
11. Bejtlich R. *The Practice of Network Security Monitoring: Understanding Incident Detection and Response*. San Francisco, CA: No Starch Press, 2013.
12. Valeriano B, Maness RC. *Cyber War versus Cyber Realities: Cyber Conflict in the International System*. New York: Oxford University Press, 2015.
13. Demchak CC. *Wars of Disruption and Resilience: Cybered Conflict, Power, and National Security*. Athens, GA: University of Georgia Press, 2011.
14. Lynn III WJ. Defending a new domain. *Foreign Affairs* 2010;**89**:97–108.
15. Rid T, Buchanan B. Attributing cyber attacks. *J Strat Stud* 2015;**38**:4–37.
16. Gartzke E, Lindsay JR. Weaving tangled webs: offense, defense, and deception in cyberspace. *Secur Stud* 2015;**24**:316–348.
17. Rid T. Cyber war will not take place. *J Strat Stud* 2012;**35**:5–32.
18. Valeriano B, Maness R. The dynamics of cyber conflict between rival antagonists, 2001-2011. *J Peace Res* 2014;**51**:347–60.
19. Lindsay JR. Stuxnet and the limits of cyber warfare. *Secur Stud* 2013;**22**:365–404.
20. Weimann G. Cyberterrorism: the sum of all fears? *Stud Confl Terror* 2005;**28**:129–49.
21. Libicki MC. *Conquest in Cyberspace: National Security and Information Warfare*. New York, NY: Cambridge University Press, 2007.
22. Dunn Cavelty M. Cyber-terror—looming threat or phantom menace? The framing of the US cyber-threat debate. *J Inf Technol Polit* 2008;**4**:19–36.
23. Cornish P, Livingstone D, Clemente D, *et al. On Cyber Warfare*. London: Chatham House, 2010. http://www.chathamhouse.org.uk/publications/papers/view/-/id/967/.
24. Liff AP. Cyberwar: a new 'absolute weapon'? The proliferation of cyberwarfare capabilities and interstate war. *J Strat Stud* 2012;**35**:401–428.
25. Betz D. Cyberpower in strategic affairs: neither unthinkable nor blessed. *J Strat Stud* 2012;**35**:689–711.
26. Rid T. *Cyber War will not take Place*. London: Hurst, 2013.
27. Lindsay JR. The impact of China on cybersecurity: fiction and friction. *Int Secur* 2014;**39**:7–47.
28. Benson DC. Why the internet is not increasing terrorism. *Secur Stud* 2014;**23**:293–328.
29. Arquilla J and Ronfeldt D (eds). *In Athena's Camp: Preparing for Conflict in the Information Age*. Santa Monica, CA: RAND Corporation, 1997.
30. Eriksson J, Giacomello G. The Information revolution, security, and international relations: (ir) relevant theory? *Int Polit Sci Rev Rev Int Sci Polit* 2006;**27**:221–44.
31. Choucri N. *Cyberpolitics in International Relations*. Cambridge, MA: MIT Press, 2012.
32. Kello L. The meaning of the Cyber Revolution: perils to theory and statecraft. *Int Secur* 2013;**38**:7–40.
33. Clausewitz C. von. *On War*. Princeton, NJ: Princeton University Press, 1976.
34. Schelling TC. *Arms and Influence: With a New Preface and Afterword*. New Haven, CT: Yale University Press, 2008.
35. Fearon JD. Rationalist explanations for war. *Int Organ* 1995;**49**:379–414.
36. Powell R. *In the Shadow of Power: States and Strategies in International Politics*. Princeton, NJ: Princeton University Press, 1999.

37. Glaser CL. *Rational Theory of International Politics: the Logic of Competition and Cooperation*. Princeton, NJ: Princeton University Press, 2010.
38. Waltz KN. *Theory of International Politics*. Reading, MA: Addison-Wesley Pub. Co., 1979.
39. Keegan J. *A History of Warfare*. New York: Alfred A. Knopf, 1993.
40. Gartzke E. War is in the error term. *Int Organ* 1999;**53**:567–87.
41. Kaplow JM, Gartzke E. Knowing unknowns: the effect of uncertainty in interstate conflict. 2015.
42. Slantchev BL. Feigning weakness. *Int Organ* 2010;**64**:357–88.
43. Blainey G. *Causes of War*, 3rd edn. New York: Simon and Schuster, 1988.
44. Slantchev BL, Tarar A. Mutual optimism as a rationalist explanation of war. *Am J Polit Sci* 2011;**55**:135–48.
45. Powell R. The inefficient use of power: costly conflict with complete information. *Am Polit Sci Rev* 2004;**98**:231–41.
46. Powell R. War as a commitment problem. *Int Organ* 2006;**60**:169–203.
47. Leventoğlu B, Slantchev BL. The armed peace: a punctuated equilibrium theory of war. *Am J Polit Sci* 2007;**51**:755–71.
48. Junio TJ. How probable is cyber war? Bringing IR theory back in to the cyber conflict debate. *J Strat Stud* 2013;**36**:125–33.
49. Keohane RO. *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton, NJ: Princeton University Press, 1984.
50. Schelling TC. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press, 1980.
51. Jervis R. Cooperation under the security dilemma. *World Polit* 1978;**30**:167–214.
52. Glaser CL. Realists as optimists: cooperation as self-help. *Int Secur* 1994;**19**:50–90.
53. Lieber K. *Cyber Analogies*, Goldman EO, Arquilla J (eds). Monterey, CA: Naval Postgraduate School, 2014, 96–107. http://hdl.handle.net/10945/40037.
54. Goldsmith J. *Cybersecurity Treaties: A Skeptical View*. Hoover Institution, 2011.
55. Gompert DC, Libicki M. Cyber warfare and Sino-American crisis instability. *Survival* 2014;**56**:7–22.
56. Lindsay JR, Gartzke E. *The Power to Hurt*, Greenhill KM, Krause PJP (eds) (Under review).
57. Inkster N. Cyber attacks in La-La Land. *Survival* 2015;**57**:105–16.
58. Haggard S, Lindsay JR. North Korea and the Sony Hack: exporting instability through cyberspace. *East West Cent Asia Pacific Issues* 2015.
59. Andres RB. Inverted-militarized-diplomacy: how states bargain with cyber weapons. *Georget J Int Aff* 2014;119–29.
60. Herley C. When does targeting make sense for an attacker? *IEEE Secur Priv* 2013;**11**:89–92.
61. Yuill J, Denning DE, Feer F. Using deception to hide things from hackers: processes, principles, and techniques. *J Inf Warf* 2006;**5**:26–40.
62. Heckman KE, *et al*. Active cyber defense with denial and deception: a cyber-wargame experiment. *Comput Secur* 2013;**37**:72–77.
63. Denning DE. Barriers to entry: are they lower for cyber warfare? *IO J* 2009. http://hdl.handle.net/10945/37162.
64. Friedman AA, Mack-Crane A, Hammond RA. *Cyber-enabled Competitive Data Theft: A Framework for Modeling Long-Run Cybersecurity Consequences*. Washington, DC: Brookings Institution, 2013.
65. Lindsay JR, Cheung TM. *China and Cybersecurity: Espionage, Strategy, and Politics in the Digital Domain*, Lindsay JR, Cheung TM, Reveron DS (eds). New York: Oxford University Press, 2015.
66. Libicki MC. The specter of non-obvious warfare. *Strat Stud Q* 2012;**6**:88–101.
67. Boebert WE. *Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for U.S. Policy*, National Research Council (ed.). Washington, DC: National Academies Press, 2010, 41–52.
68. Clark DD, Landau S. *Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for U.S. Policy*, National Research Council (ed.). Washington, DC: National Academies Press, 2010, 25–40.
69. Morgan PM. *Deterrence Now*. New York: Cambridge University Press, 2003.
70. Wagner RH. Bargaining and War. *Am J Polit Sci* 2000;**44**:469–84.
71. Pillar PR. *Negotiating Peace: War Termination as a Bargaining Process*. Princeton, NJ: Princeton University Press, 1983.
72. Kalyvas SN. *The Logic of Violence in Civil War*. New York: Cambridge University Press, 2006.
73. Fearon JD. Signaling foreign policy interests: tying hands versus sinking costs. *J Confl Resolut* 1997;**41**:68–90.
74. Snyder GH. *The Balance of Power*, Seabury P (ed.). San Francisco, CA: Chandler, 1965.
75. Alexander D. U.S. reserves right to meet cyber attack with force. *Reuters*, 2011.
76. Winner L. *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought*. MIT Press, 1977.
77. *Does Technology Drive History?: The Dilemma of Technological Determinism*. MIT Press, 1994.
78. Nye DE. *Technology Matters: Questions to Live with*. Cambridge, MA: MIT Press, 2006.
79. Goldsmith J. Disconcerting U.S. Cyber Deterrence Troubles Continue. *Lawfare* 2015. https://www.lawfareblog.com/disconcerting-us-cyber-deterrence-troubles-continue.
80. Obama BH. Remarks by the President to the Business Roundtable, 2015. https://www.whitehouse.gov/the-press-office/(2015)/09/16/remarks-president-business-roundtable.
81. Clapper JR. *Statement for the Record: Worldwide Cyber Threats*, 2015. http://www.dni.gov/index.php/newsroom/testimonies/209-congressional-testimonies-2015/1251-dni-clapper-statement-for-the-record,-worldwide-cyber-threats-before-the-house-permanent-select-committee-on-intelligence.
82. Dilanian K. Intelligence chief: little penalty for cyberattacks. *Associated Press*, 2015. http://bigstory.ap.org/article/4ec29ca2b41241-c1a9fd75b3d5503c3f/intelligence-chief-little-penalty-cyberattacks.
83. Slantchev BL. *Military Threats: The Costs of Coercion and the Price of Peace*. New York: Cambridge University Press, 2011.