JOURNAL OF
CYBERSECURITY

Journal of Cybersecurity, 2022, 1–16
https://doi.org/10.1093/cybsec/tyac007
Research paper

Research paper

The illogic of plausible deniability: why proxy conflict in cyberspace may no longer pay

Justin Key Canfil[*,†]

Belfer Center for Science and International Affairs at the Harvard Kennedy School, Cambridge, MA 02138, US

[*]Correspondence address. 79 John F. Kennedy St, Cambridge, MA, US; Tel: +1 617 495 1155; Fax: +1 617 495 8963; E-mail: jcanfil@hks.harvard.edu
[†]Postdoctoral Scholar.

Received 22 November 2021; revised 11 March 2022; accepted 1 April 2022

Abstract

Cyber proxies—whether mercenaries, patriotic zealots, pranksters, or simply allies of convenience—are thought to be widespread. By outsourcing to proxies, this logic goes, a host government can plausibly deny its involvement in operations that advance its military and foreign policy aims. This presents central challenge to empirical researchers. If the value of proxies derives from their deniability, this same quality should mean that *implausibly* deniable types—the types sponsors supposedly wish to avoid—receive disproportionate attention in data and discourse. Accordingly, proxy activity appears to have flagged across several widely used datasets, depending on how the data are parsed. Do proxies still pay? A formal model is used to hypothesize about how new norms of attribution (specifically, the willingness of victims to make accusations on the basis of circumstantial evidence) can encourage capable states to *insource* more than they *outsource*. In the model, victims have the power to decide whether denials are plausible. “Usual suspects” who learn that they will take the heat regardless have fewer incentives to rely on proxies. Empirical evidence on insourcing patterns offers backdoor support for this proposition. The findings should decrease our confidence in plausible deniability as a logic for why states outsource to proxies. The paper joins an emerging body of research that has questioned the role of plausible deniability in covert action, including cyber conflict.

Key words: cyber conflict, cyber proxies, state-sponsored, attribution, cyber norms, plausible deniability

Former president Dwight D. Eisenhower once remarked that proxy conflict is “the cheapest insurance in the world” [1]. But why would a state outsource something it can do for itself? In the classical framework, states outsource to non-state proxies in order to obscure their involvement [2–4]. By refraining from direct action, all the while tacitly permitting or actively supporting proxy activity, host governments are able to enjoy foreign policy or military gains without admitting to culpability. Host states that do this successfully elude the usual risks from using force against a capable adversary, ranging from sanctions and reputational damage [5] to armed escalation [1,6,7]. This logic is widely assumed to hold in cyberspace, as well [8,9]. And yet, if cyber proxies offer such clear advantages, it is puzzling why this strategy has not universally proliferated, especially among states that are host to highly capable and ideologically aligned hacker collectives.

How have global patterns in the use of cyber proxies changed, if at all? This paper supposes that proxy conflict in cyberspace may no longer pay—at least not for the reason scholars ordinarily assume, plausible deniability. Despite its status as a widely assumed causal mechanism, deniability for proxies may not very plausible, after all. Victims set the evidentiary threshold for themselves. In fact, victims are free to make whatever allegations they like. Proxies sometimes even out themselves to claim credit [10], undermining any remaining veneer of secrecy. Nor in many cases do alleged sponsors even bother to deny victims’ accusations, e.g. because there is signaling value in implicating oneself [11]. Deniability is only plausible to the extent it discourages retaliatory action. If victims are increasingly willing to respond on the basis of imperfect attribution [12,13], attackers cannot hope to gain much political cover by outsourcing to proxies. Without plausible deniability, proxies may lose their luster.

Scholars of cyber conflict would do well to take seriously this proposition. Writing on reasons for covert action in physical domains, Cormac and Aldrich [14] argue that "even in its supposed heyday, the concept [of plausible deniability] was deeply problematic. Changes in technology and the media, combined with the rise of special forces and private military companies, give it even less credibility today." These include advances in attribution technology, attribution from the private sector, media coverage, and the proliferation of in-house government cyber capabilities. Similarly, Lin-Greenberg and Milonopoulos [15] and Vaynman [16] have argued that open-source surveillance and monitoring technologies make it difficult for governments to hide their illicit activities in any domain. It stands to reason that this logic would hold for cyberspace as well. For instance, private cyber analysis firms have proliferated in number, and face fewer political barriers than states in making independent attributions. As Cormac and Aldrich [14] explain, "we live in an era of implausible deniability."

In this paper, a formal model is used to generate hypotheses about why states might choose insourcing over outsourcing, even when proxies are both highly capable and perfectly committed to carrying out their orders. The model relaxes a standard assumption that deniability is solely a function of the sponsor's level of investment, instead considering the target state's willingness to cast blame even when evidence connecting the sponsor to its proxy is circumstantial. The minimization of principal–agent problems in the model is helpful an as extreme test of the theory that proxies might not pay for external reasons. The theoretical predictions are supported by new empirical data. Deniability is least plausible when there is clear evidence that a sponsoring state conducted an operation directly, insourcing to its own personnel. I find that public US reprisals are associated with more insourcing, as opposed to more outsourcing, in subsequent operations. These findings contribute to an emerging body of research that has questioned the logic of plausible deniability in covert action, including cyber conflict [10,11,14].

The proposition that proxies may no longer pay is intended to be suggestive, *not* definitive. Despite the confirmatory evidence, the paper stops short of asserting that proxy conflict is in fact known to be declining. Instead, the argument serves to remind scholars to question their underlying assumptions about why states outsource, given that alternative explanations make predictions with observationally equivalent empirical implications. We can, however, measure and test the converse—insourcing. The findings show that insourcing is on the rise, especially in response to finger-pointing by powerful victims. Because victims are increasingly willing and able to blame, and because the obviousness of insourcing shatters any remaining plausible deniability, these findings suggest that outsourcing may not be as *en vogue* as is widely assumed for aggressors who desire political cover.

That is not to say that proxies *never* paid, or do not pay still under certain circumstances. To date, some 400 000 multinational hackers have reportedly volunteered to aid the besieged Ukrainian government against Russia in the ongoing 2022 war [17]. Ukrainian officials have openly welcomed their assistance. Ukraine's Minister of Digital Transformation even tweeted an open invitation to join his "IT army" and to "fight on the cyber front" [18]. By doing so, of course, Ukraine cannot plausibly deny a relationship with its proxies. Nor would it presumably care to, given the state of war. Proxies may still sometimes pay, but if or when they do, it must be for reasons *other* than political cover.

Cyber Proxies: What We Know

Whether driven by ideological zeal, nationalism, or money, non-state actors have long been a persistent phenomenon in international conflict, including cyber conflict.[1] For instance, in 2001, Chinese hacktivists, encouraged by the state, targeted US government websites in retaliation for the EP-3 incident [24–27]. Taiwanese networks were also targeted during a period of growing cross-Strait tensions that same year [28]. Even the world's most notorious cyber conflict "wake-up call" [29], the attack on Estonia in 2007, was putatively facilitated by a pro-Russian youth organization [24,30]. Such observations undoubtedly contributed to predictions that cyberspace would profoundly level the playing field between nation states and non-state actors [31].

But states have not disappeared as a central actor, even in cyberspace. States which are host to non-state hackers can often co-opt these actors in order to advance operational objectives [32]. As Akoto [33] explains, "[cyber] operations have proliferated... [and] states are increasingly outsourcing them to non-state actors"—also known as "cyber proxies."[2] I use the term "sponsor" to refer to a sovereign power that relies on non-state actors to complete some operational cyber objective; to the non-state actor as that sponsor's "proxy;" and to this relationship as "outsourcing," as opposed to "insourcing:" centrally coordinated, direct action by the sponsor.

This of course describes a very general relationship. Scholars recognize a range of connections states might have with non-state actors. Healey [34] identifies up to 10 types of arrangements, ranging from abeyance or disavowal at one extreme to command centralization at the other. Maurer [35] distills this into three core managerial strategies: delegating, funding, and tacitly permitting.[3] More recent work by Egloff [32] refines this further. Such frameworks are useful and come with one unsurprising takeaway: the greater the level and intensity of state support, the greater the objective certainty that the state is formally responsible [36].

Detection and attribution in cyberspace are challenging, but not impossible. Because of this, it is widely believed that the use of a proxy bestows sponsoring states with an added layer of plausible deniability [8,35,37–42]. Cyber proxies by definition operate under a different command structure and may work physically apart from regular uniformed personnel. Even if a perpetrator can be tracked down, organizational separation between that perpetrator and the principal to whom she reports can make it difficult for victims to know the principal's involvement. Plausible deniability helps distance sponsoring states from the risk of attribution. Victim states may find

[1] Long ago common in naval warfare, proxy conflict once again became prevalent in the latter half of the 20th century as armed attack and conquest became less accepted as legitimate instruments of statecraft [19]. Scholars have found that a majority of rebel groups since 1945 receive tacit state support [20], and this support has made these groups more powerful and influential [21]. They are observed most often between great power dyads [6,22,23], presumably because the risks associated with direct action are higher when targets can strike back.

[2] Proxies "may or may not be part of state security and intelligence agencies and may be criminal syndicates or private cybersecurity companies... Some groups may start off as private hacker collectives and are then absorbed into state security agencies or vice versa" [33].

[3] Maurer [35] writes that state sponsors can actively fund or otherwise support non-state cyber actors with the expectation that these actors will accomplish some military, intelligence, or foreign policy objective on behalf of the state (what Maurer terms "orchestrating"). Alternatively, sponsors can simply permit offensive operations by non-state actors by turning a blind eye ("sanctioning"). At the opposite end, a sponsor can establish effective in-house command and control over the group ("delegating").

it hard to exercise prosecutorial jurisdiction over individual hackers, especially where extradition agreements are lacking, so law enforcement cannot lean on the culprits to implicate their state sponsors [43]. If true, then it stands to reason that states can use their proxies to secure operational objectives at low cost and low risk of attribution.

Some non-state hackers cooperate with states for ideological reasons, not just money or glory. So-called "patriotic hackers" first appeared as early as 1998 [44]. Among cyber proxies, patriotic hackers should offer their sponsors the best of both worlds. Would-be state sponsors face a classic principal–agent dilemma: by empowering non-state hackers, the state reaps the benefits of an operation while minimizing its exposure if the operation is traced to the perpetrators. But non-state hackers may have their own agenda. States that turn a blind eye run the risk of operational deviations, and states that provision their proxies with resources and material support may face a "Promethean dilemma" if those proxies later use their newfound capabilities against their sponsors [8]. Because patriotic hackers support state objectives for ideological reasons, the sponsoring state need not rely on side payments [45] to keep them in line. Without a money trail leading back to the sponsoring state, the usual principal–agent risks are minimized: the state can be assured of its proxy's fidelity, and plausible deniability is strengthened.

This logic has led many scholars to assume that reliance on cyber proxies must be widespread. Russia [35,46], China [25,26], Iran [40], and North Korea [47]—all of which possess their own powerful cyber capabilities—are usually regarded as the most complicit. Standard accounts rarely note how the outsourcing model has been at various stages endorsed in countries like Japan, India, and Estonia [45,48], and (allegedly) employed by several Western countries [38,42, 49–51]. On the other hand, scholars acknowledge that proxies are not ubiquitous. The US has traditionally kept its non-state hacker community on a tight leash by discouraging operational participation [35,46].[4] Between 9/11 and the start of the Iraq War, e.g. when US nationalism soared, the US National Infrastructure Protection Center (NIPC) posted a notice explicitly warning overzealous Americans not to hack Iraqi targets [53]. On the surface this might seem surprising, since the US shown little reluctance to engage in proxy warfare in other domains.

So, just how prevalent are cyber proxies? How can we know? When making inferences about global patterns, researchers must rely on large-*n* datasets. The data are based on public reports and involve multiple hurdles—detection by the victim, admission and disclosure, sufficient evidence, media attention, and attribution. As Akoto [33] explains, these data limitations have "crippled efforts to study... cyber proxies." Only in cases where the target detects an intrusion, admits to it, and alleges the involvement of a foreign state sponsor will that case will show up in open-source cyber conflict datasets.

The third point—allegations—is key. The relationship between a sponsor and its proxies must be inferred by the target or other observers. Especially bold (or frustrated) victim states might overstate their level of confidence about who the culprit is. Cautious victims might decline to make allegations with any certainty at all. What this means is that the criteria for inclusion in the universe of observed cases is a function of victim transparency and observer beliefs.

For researchers interested in studying cyber proxies, this raises a troubling prospect: the dominant explanation for why states outsource to proxies is plausible deniability, but scholars can only observe cases in which *denials were not considered plausible enough* to prevent attribution. Scholars of cyber conflict have done excellent work to circumvent data problems in other areas [54–56], but for this reason the study of cyber proxies faces a special set of challenges.

Observational equivalence in existing data

The logic of the problem is thus. Assume outsourcing bestows plausible deniability by making attribution harder, and that plausible deniability is something states desire. If outsourcing is on the rise as is commonly assumed, it follows that we should actually record *fewer* cases in the data over time, not more, because attributions will seem incredible. On the other hand, if outsourcing does not convey much plausible deniability, or if states have more interest in operational reliability than plausible deniability, we would also record fewer cases. This would be for precisely the opposite reason: because outsourcing is no longer an unattractive strategy. There is no direct way to adjudicate between these processes with the available open-source data.

To illustrate this problem of observational equivalence, I compare observations from three datasets: Valeriano [56]'s Dyadic Cyber Incident and Dispute Dataset 1.5 ("VJM");[5] the Council on Foreign Relations' Cyber Operations Tracker ("CFR") [57];[6] and Akoto [33]'s dataset on cyber proxy onset.[7] CFR [57] and VJM [56] track incidents in which state involvement was suspected, but neither differentiates between insourced and outsourced operations.[8]

One reasonable expectation might be that states are likelier to delegate certain tasks to proxies—e.g. less sophisticated or less sensitive operations. We can subset these datasets to include only attacks *other than* espionage and Advanced Persistent Threats (APTs), respectively, to infer how much of this activity is driven by nation state operatives versus their proxies.[9] A third dataset by Akoto

[4] There are notable exceptions. In November 2021, e.g. a hacker group called "Hooshyarane Vatan" doxxed Iran's airline Mahan Air. The group explained on Twitter that the hack was aimed at bringing about "justice for the Ahwaz," a persecuted Iranian minority group. Was this a hacktivist incident or a proxy attack? On the one hand, according to *The Jerusalem Post*, Ahwaz hackers "carry out various cyber and physical operations against the [Iranian] regime periodically," and are enabled by "technological and other assistance from Israel [and] the US" [52].

[5] https://drryanmaness.wixsite.com/cyberconflict/cyber-conflict-dataset

[6] https://www.cfr.org/cyber-operations/

[7] https://journals.sagepub.com/doi/suppl/10.1177/07388942211051264

[8] In the CFR dataset, a "PLA [China's People's Liberation Army] unit or a tech company that works for MSS [the Chinese Ministry of State Security]" would be coded identically. For instance, the data include an offensive campaign by US Cyber Command against Russia's Internet Research Agency in 2018, which most certainly was not outsourced to a non-governmental threat actor. Thanks to Adam Segal and Connor Fairman for clarification on these points by email.

[9] Subsets are based on the following rationale. First, the sponsor may not trust its non-state agents to handle sensitive intelligence. Whether an attacker shares indicators of compromise (IoC) with Advanced Persistent Threat (APT) may also be relevant. APTs are sophisticated, highly organized, recurrent threats commonly equated with known nation state actors. Similarly, less sophisticated proxies, such as loosely organized patriotic hacker groups, may be more likely to employ nuisance tactics such as distributed denial-of-service (DDoS) attacks, website defacements, or doxxing campaigns.

[33] contains an extensive survey of known proxy groups.[10] However, it only tracks proxy relationship onset.[11] As a strictly cumulative measure, it cannot be used to measure whether outsourcing has declined. Instead, we can use it to observe the rate at which states have continued to outsource to new groups in recent years.

Figure 1 compares observations from all three datasets. The *Y*-axis in the first two rows depict the number of reported cyber incidents in the CFR/VJM data. The the third row maps the total number of proxy relationships in the Akoto dataset over time. Loess curves (95% confidence intervals) measure the change in *Y*-values over time, sorted by country faction.[12] Next, contrast the left and right columns for each row. Left columns in the top two rows plot the full universe of reported operations (CFR; VJM) and total number of proxy relationships in the international system (Akoto). Conversely, the right columns subset that data on the correlates of proxy activity (i.e. operations other than espionage and APTs). Meanwhile, on the third row, the right column plots the number of newly added proxies.

We can see that although incidents have become more frequent overall according to CFR (top row), this activity is driven almost entirely by Russian and Chinese espionage. Likewise, the VJM data depicts an apparent decrease in activity other than APTs, especially after 2013. Differences in the Akoto data, which explicitly tracks proxy activity, are most striking. At first glance, the left column of the third row would seem to indicate that US adversaries have drastically increased the number of groups to whom they outsource. However, keep in mind this is strictly a cumulative measure, and does not track when groups disband or are abandoned or absorbed by their sponsors. The right column tracks the onset of new relationships. It becomes apparent that the *rate* of outsourcing flags after 2010.

These trends could tell one of two mutually exclusive stories. At first glance, it might seem that outsourcing is actually less common now than is usually assumed. Alternatively, Akoto [33] explains that in order for the onset of a proxy relationship to appear in the dataset, "there must be *strong evidence that a group is acting on behalf of*, at the behest of or with the active support of the government" (emphasis mine). Akoto [33] continues: "It is insufficient for a group to be considered a proxy simply because it is ideologically aligned with the government, shares a common enemy or does not oppose the government... [as long as the group is] somewhat independent of the state."[13] Data collection is scoped on proxies which are already *known* to work for the state. Inclusion in the dataset is therefore inversely related the plausibility of a sponsor's deniability.

What this means is that the large-*n* information we do have consists of cases in which denials were implausible. If proxies really *do* convey plausible deniability, then they are less likely than insourced operations to appear in datasets that treat nation states as the unit of analysis. Assuming plausible deniability is in fact the main reason states outsource, the data systematically exclude exactly the types of proxies to whom states have any interest in delegating.

Direct measurements of outsourcing in available data cannot distinguish between these competing accounts. On the one hand, outsourcing could be more common because sponsoring states are outsourcing to proxies who are more deniable. If states can in fact use proxies to achieve plausible deniability, the dearth of empirical evidence to attest to the fact is only natural, because academics would be among the last to become aware of the true relationship. On the other hand, outsourcing could really be less common now because sponsoring states are learning that their deniability is not as plausible as they had hoped. After all, if academics can code your relationship, it is unlikely to fool a sophisticated adversary.

The Illogic of Plausible Deniability

It has long been assumed that proxy conflict is attractive precisely because it offers state sponsors plausible deniability. This section theorizes about why that might no longer be true.

In the early days of cyber conflict, attack attribution [59,60] was widely regarded as its "most difficult problem" [61]. The barriers to attribution prevented victims from identifying and holding aggressor accountable. It more recent years, however, the attribution problem has come to be seen as less problematic [62]. Attribution may be onerous and complicated, but it is technically possible [12,13,63]. Victims typically work to piece together an idea of the likeliest culprits based on indicators of compromise (IOCs), an accumulation of technical clues and circumstantial patterns [64].

Even if victims are able to trace the source of an attack, knowing who the people behind the terminal are and who they work for is a far more challenging set of questions. First, private sector targets are often the gatekeepers of forensic evidence and may have countervailing incentives not to disclose [33]. States are often sensitive about divulging how they know what they know [65]. Assuming a victim state is willing to broadcast the evidence it collected, the burden of proof to implicate a sponsoring state is prohibitively high as a matter of international law. Though the law is not completely settled, jurists usually interpret it to mean that only sponsors who take on a direct command role can be held responsible for the actions of their proxies [36,66].[14]

[10] Akoto [33] takes a very broad view of what constitutes a "proxy" operator, including private software providers and defense contractors such as Raytheon and Lockheed Martin.

[11] In particular, the Akoto dataset does not contain information about whether the state partnered with a proxy only temporarily; a proxy group was disbanded; or a longstanding relationship was severed.

[12] Solid lines ("Blue," named for *blue space*) plot the trend for the US, EU, Israel, and other "like-minded countries" in the Western coalition. Dashed lines ("Red," for *red space*) include "the usual suspects"—Russia, China, Iran, North Korea, and Shanghai Cooperation Organization members. Dotted lines ("Gray," for *gray space*) include all other states, such as the non-aligned countries, though these are severely underrepresented in all three datasets.

[13] Akoto [33] also explicitly excludes "flash" groups—groups that emerge organically to conduct operations. This would seem to eliminate virtually all patriotic hackers, the type of proxy with the *most* deniability (see [34]).

[14] The International Law Commission (ILC)'s 2001 Draft Articles is perhaps the clearest articulation of the law on state responsibility (though often cited, it is actually a nonbinding source). In the ILC's view, a host state can only be held responsible if it directly conducts or oversees the offending operation at the operational level (see International Law Commission (ILC) Draft Articles on State Responsibility, 2001, http://legal.un.org/ilc/ texts/instruments/english/commentaries/9_6_2001.pdf). International courts have held variously that the relationship must be one of "effective control" (Nicaragua v. United States), "overall control" plus the commission of specific acts (Prosecutor v. Dusko Tadic (1999)), or "complete dependence" (Bosnia and Herzegovina v. Serbia and Montenegro (1996)) – all very strict standards that elide the possibility of measures short of command or operational involvement. And in more than 4,700 claims proceedings, all but one tribunal (Dames & Moore v. Iran) found that the burden of proof for wrongdoing rests with the defender (see Iran Claims Tribunal, https://www.state.gov/iran-u-s-claims-tribunal/).

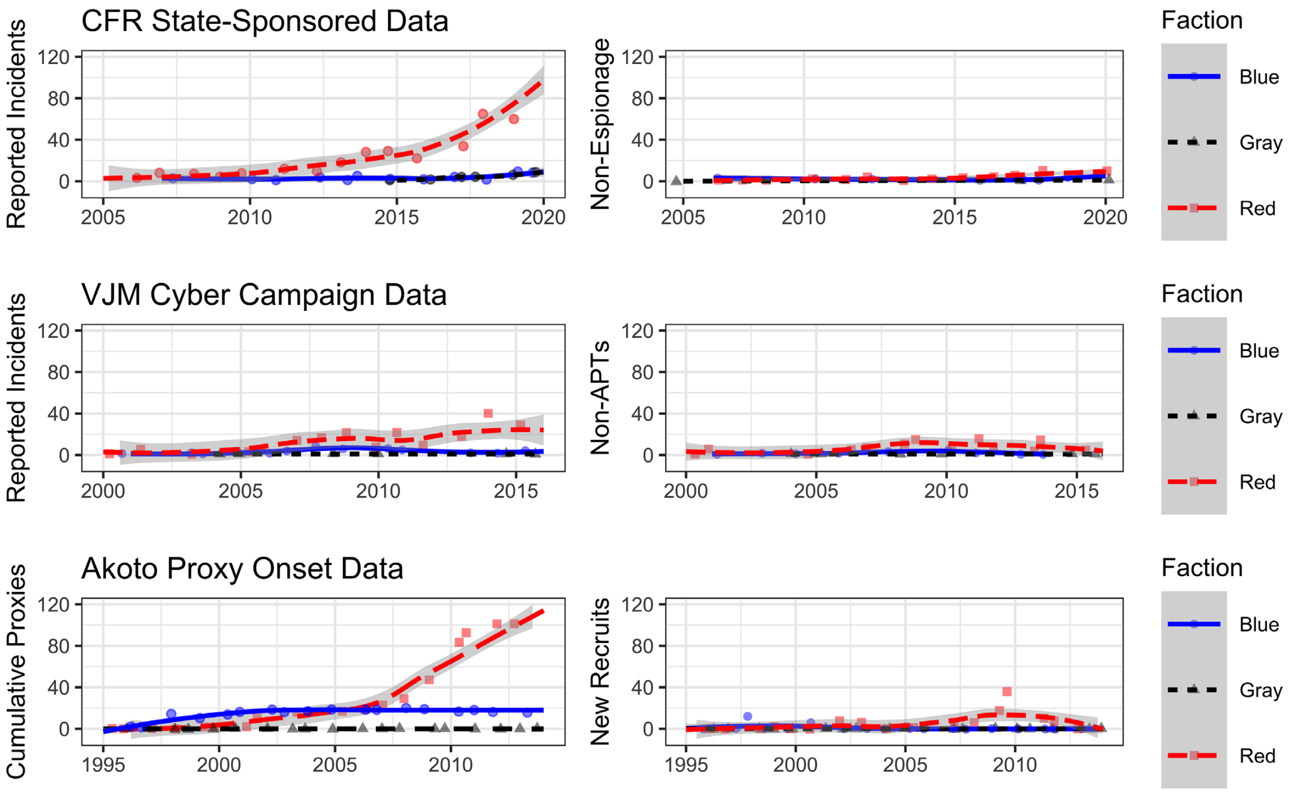

Figure 1: Comparisons of trends over time from three datasets [33,57,58]. The leftmost column plots the full set of observations by sponsoring faction. The rightmost column plots correlates of proxy support. Loess curves (95% confidence intervals) plot trends by faction. Factions are organized around US allies (blue), adversaries (red), and other states (gray). After subsetting on likely proxy activity (right), claims that proxies are widespread seem much weaker, especially in the last decade.

These barriers are what make outsourcing so attractive, according to conventional wisdom [35]. But assuming cyber proxies convey plausible deniability for their sponsors, and sponsors find plausible deniability desirable, why is reliance on proxies not universal [33,67]? In the US case, scholars often point to philosophical disagreements between Washington and Silicon Valley (e.g. [35]). Others have argued that authoritarian countries and democracies face disparate accountability barriers [33,68]. A debate exists as to the real prevalence of proxies among authoritarian countries, as well (see [69]). Russia and China have since 2016 exhibited an increased interest in more tightly controlled cyber operations. Both countries have adopted stronger and more centralized domestic cyber institutions [55], and in 2016 Segal [70] described plans to adopt a model reminiscent of US Cyber Command.

From 1994 to 2003, the Chinese government was suspected of tacitly permitting its proxies to attack foreign targets, and from 2003 to 2013 the state appeared to provide active support [28]. Despite rising online nationalism in China [71] and an army of cyber privateers that once purportedly numbered almost 200 000 [24], however, Beijing has since tightened its enforcement of internal cybersecurity laws. For example, Hang [72] catalogues 11 patriotic hacking episodes alleged to have been sponsored by China. Yet, all but four occurred prior to 2001 and none after 2010. The Honker's Union (中国红客), reportedly began as a Chinese patriotic hacker collective in the 2000s [see 73]. In 2011, its leader warned his colleagues that they "probably won't" be permitted by the government to continue attacking foreign targets [74]. Since then, the Chinese government has relied more on specialized units at the Ministry of State Security [35,70]. Attacks originating from that country reveal a more direct government hand [75].[15] If proxies pay for the reasons we usually suspect, then this behavior is puzzling.

Rarely do cyber attacks leave smoking gun evidence about who was involved. Attribution is already hard, so plausible deniability should make for an additional hurdle. Yet, victims can and do increasingly point the finger, even when attributions are imperfect [12,13,76,77]. *Cui bono* ("who benefits?") tests are themselves commonly regarded by decisionmakers as a kind of evidence. A victim may still suspect its adversaries of being behind an attack perpetrated by ostensibly non-state hackers. In such a case, the target state might privately or publicly respond through an act of retorsion, naming and shaming, sanctioning, or launching a proportional cyber response. Since 2014, the US and other powerful states have even begun to publicly attribute cyber incidents to specific individuals that they say were working on behalf of sponsoring governments. Sometimes, these are based on extensive evidence. In other cases, detailed evidence is not provided.

In particular, the propensity of the US Department of Justice (DOJ) to indict foreign hackers has increased in recent years, despite the difficulty of trying these defendants in the US courts [43]. In 2014, the US DOJ famously indicted five Chinese People's Liberation

[15] For another example, see a 2019 CERT-EU memo: https://media.cert.europa.eu/static/MEMO/2019/TLP-WHITE-CERT-EU-MEMO-190729-1.pdf.

Army (PLA) officers for allegedly "hacking under the shadow of their country's flag" in order to escape punishment [78]. Though an indictment under the US criminal law signaled the government's view that the suspects were individuals, the attribution to China was explicit. Comparing this 2014 indictment to the Equifax hack 3 years later, observe that DOJ named 30 accomplices [79]—a 6-fold uptick in the number of implicated individuals. The US government has continued to aggressively indict foreign hackers it considers to be nation-state proxies (e.g. [80–84]). Speaking on behalf of the US Government, former US Acting Deputy Attorney General John Carlin argued that proxies "are not immune from the law" and that the US "will hold state sponsored cyber thieves accountable" [78].

It is worth examining how, practically speaking, the US DOJ is able to piece together an argument that cyber incidents are indeed "state-sponsored."[16] In a separate piece, Carlin [43] explains that the Criminal and National Security Divisions at DOJ "increasingly find [them]selves working cases jointly (or at least more actively supporting each other's cases)," including by embedding special agents from Counterintelligence with the FBI's Cyber Division. Enabled by post-9/11 changes to the structure of the intelligence community, these interagency and public-private collaborations—in Carlin's own words—were motivated by the early cyber proxy landscape.[17]

This level of cooperation gives the government specially enhanced attribution capabilities. Carlin [43] describes the idea that the US has an attribution problem as "antiquated." As former National Security Agency General Counsel, Stewart Baker stated in 2015, "we can [now] know who our attackers are" (Congressional testimony, quoted in [43]). Victims may be more willing to attribute their attackers when an attack has large-scale and high-profile consequences, even if evidence is weak [13].

Especially for these types of cases, DOJ has a strong motivation to attribute a state sponsor. DOJ depends on court-issued warrants for its investigatory powers, and warrants are often obtained by naming a nation state. As Carlin [43] admits, nation state attribution is actually crucial in the Foreign Intelligence Surveillance Act (FISA) process: "the government must demonstrate, among other things, that the 'target... is a foreign power or an agent of a foreign power'" if it hopes to obtain a FISA warrant. Presumably this creates pressure for DOJ to implicate uncooperative states who are host to cyber attackers, whether or not the state-proxy evidence is rock solid.

Indictments still require a degree of legally admissible evidence, even if the governments understands the ability to actually convict on this evidentiary basis will never be tested. But even short of formal indictments, the US began acting on its suspicions when it initiated its new operating concept, persistent engagement [84–86]. To my knowledge, there is no institutional requirement to show that the target of a "hunt forward" operation has legally demonstrable links to a host government. Would-be sponsors must certainly be aware that outsourcing does not automatically bestow political impunity. At best, outsourcing might only convey international legal protection in an environment where international law is already permissive; i.e. for operations below the level of an armed attack (see [87]). Where once victim states like the US exhibited "restrained responses" [88], they now seem more willing to punish sponsors for the activities of their proxies.[18]

After all, why would not victims overplay rather than underplay their hand? If international relations were a courtroom, then proxies might not implicate their sponsors beyond a reasonable doubt. But pointing the finger in international relations is far more political. It is the victim that decides the burden of proof. Political attribution bridges the gap between technical facts and suspicions, "reducing uncertainty about who is behind an intrusion" and "creating cybersecurity 'truths'" for victims with the means to retaliate [89]. States may be learning that they can expect to be blamed even when victims cannot establish a rock solid evidentiary link. Even refusal to cooperate in stopping attacks emanating from within one's borders might be construed as tacit support. If so, advances in the technology of attribution, coupled with victim states' newfound willingness to make allegations on the basis of strong suspicion rather than hard evidence, may mean that deniability offers fewer advantages than it once did.

Raising suspicions may be enough to incur a victim's ire. As long as perpetrators can be linked with some confidence to possible sponsor—even if that confidence is not absolute—the plausible deniability assumption may not hold.

A model of outsourcing

How can we study the proposition that proxies might no longer pay? One way of challenging assumptions is to check whether one's assumptions are sensitive to reconfiguration. Formal models are the appropriate technology for checking and comparing a theory's internal validity this way. Following scholars such as Lindsay [13], Baliga *et al.* [12], Axelrod and Iliev [90], and Merlevede et al. [91], I formally model an aspect of cyber conflict. I formally model an aspect of cyber conflict. Despite the use of formal models for studying proxy relationships in other domains, this technology has rarely been applied to questions of cyber conflict and to the author's best knowledge has never before been applied to cyber proxies. Formal models are not empirical tools *per se*, but they do allow researchers to understand how data-generating processes may work "under the hood" in cases where data are questionable or unavailable.

Every theory is based on core assumptions. The literature on cyber proxies has long assumed three points. First, cyber proxies offer plausible deniability. Second, sponsors desire plausible deniability, and therefore, find proxies useful. Third, because of this, the use of proxies must be widespread. (Absent large-*n* data, these assumptions have been difficult to challenge.) The third point has been demonstrated through anecdotes and case studies, but it has been difficult to measure changes in the global rate over time. Formal models can be used to illustrate how processes very different from the ones we assume can produce observationally equivalent, but possibly spurious, outcomes [92].

In the model, a host government faces a choice between outsourcing (turning a blind eye or actively supporting non-state actors) versus insourcing (conducting cyber operations using its own, in-house capabilities). Evidence is a function of how much a sponsor invests in the units charged with carrying out the attack. The more is invested, the more obviously the sponsor is linked to the operation. However, a key difference in the model is to relax the dependence between victims' beliefs and the level of investment. Under this framework, the sponsor's deniability is a function of its involvement *and* the target state's independent willingness to make accusations. This is a generalization of the conventional wisdom, since it does not assume that all victims have an identical standard for the burden of proof.

The model shows how a sponsor's concern that victims will retaliate on the basis of imperfect or indirect evidence can be sufficient

[16] For a detailed exposition in this journal, see [77].

[17] "The most sophisticated threats we investigate are associated with nation-state actors or their proxies" [43].

[18] Note that Kaminska [88]'s argument in this journal is that US hesitancy to respond "stem[med] from a desire to avoid risk," *not* uncertainty about the identity of the culprit.

to discourage that sponsor from outsourcing. Note that this holds even when proxies are ideologically faithful and highly competent. In the real-world, outsourcing comes with all the downsides typically associated with a principal–agent relationship [8]. The model holds these widely established principal–agent problems constant at their ideal, setting them aside so that other mechanisms can be studied under controlled conditions. For this reason, the decision to outsource can be modeled as decision-theoretic. This permits a more specialized focus on plausible deniability's role in decisions about whether to outsource.

Model setup

A host government (G) oversees a pool of non-state hackers (H). The government is interested in engaging in cyber operations, but must decide whether its offensive cyber operations should be, on balance, more centralized, or outsourced. Define this decision as $\varphi \in [0, 1]$, where $\varphi > 0$ is some positive degree of outsourcing. One can imagine that low levels of φ represent tacit permissiveness, higher levels represent encouragement, and the highest levels represent funding, support, and direction. Conversely, the government runs its own operations and discourages non-state activity at $\varphi = 0$, for instance by implementing and enforcing laws against hacking foreign targets. In the game, G's campaign payoff aggregates an arbitrarily high number of operational payoffs given its chosen degree of command centralization. For ease, Table 1 contains a guide to notation.

The government's return on investment is given as $\tilde{u}x$, where x is its aggregate level of investment. Investment determines how intrusive the attacks are, and therefore, how much information or destructive opportunity they yield. Penetration capabilities also vary by the attacker's level of sophistication, ψ_i. The probability of a successful campaign is ψ_i. I assume that individual operations are scalable, and thus the government obtains either full benefit $\tilde{u}x$ from a successful campaign ($\psi = 1$) or no benefit at all ($1 - \psi$).[19]

Consistent with default explanations, G faces principal–agent problems *vis-a-vis* H. The degree of agent drift—a function of H's operational discipline and its ideological sympathy with government objectives—is written as θ. The government's level of investment x, and thus its expected payoffs, are penalized increasing in θ. In other words, the less reliable the agent ($\theta\uparrow$), the less useful its efforts.

Finally, G gains some plausible deniability depending on its centralization strategy. The extant literature assumes that plausible deniability is a function of the level of centralization, $f(\varphi)$, and that the attacker faces penalties increasing in attack scale, x. This would be modeled as $x \times \varphi$: the less centralized, the more deniability when scale is held constant.[20] The legal standard for attribution, as stated previously, requires *effective control* over command and operations; simply permitting, encouraging, or even funding is not sufficient to trigger a target's right to self-defense. Thus, only centralized operations should impose attribution penalties under default frameworks. Yet, we know this to be false, since victim states are increasingly willing to hold host governments liable for the actions of their agents.

Departing from conventional accounts that tend to conflate different modes of attribution, this model distinguishes between *technical attribution*, denoted τ, and *political attribution*, α. Attribution τ_v is simply the probability of detection, determined by the target state's capabilities. Under the status quo law of state responsibility, even very sophisticated forensic techniques are often of little help to victims in making a case against sponsoring states for the actions of their agents. The target state must be able to show effective command oversight, and the burden of proof is on the target. This is usually not possible through technical means alone, unless the agents operate directly out of host government networks. Conversely, *political* attribution, $\alpha_v > 1$, is determined by the target state's tolerance of circumstantial or legally insufficient evidence. In other words, technical attribution reflects a state's capability, whereas political attribution represents its willingness to act on the basis of non-technical evidence.

The government's payoff can be modeled as

$$U_G(\varphi) = \psi\tilde{u}x(2 - \varphi) - \theta\varphi x - \gamma(1 - \varphi) - \tau_v c\left(\alpha_v(\varphi) + x(1 - \varphi)\right).$$

In plain terms, G's expected benefit is given by the balance of participation, $\tilde{u}(2 - \varphi)$, times the probability the campaign is successful (ψ), and the amount of material investment required (x). If the campaign is successful, G's investment yields a return of $\tilde{u}x$. G has a choice whether to outsource ($\varphi > 0$) or centralize the campaign ($\varphi = 0$). Outsourcing risks agent drift, with departure costs increasing in agent misalignment times the level of investment (θx). Centralization may also involve investment of some fixed downpayment γ. For example, the government might need to purchase buildings, reorganize cyber units, write contracts, or issue uniforms. Finally, G incurs some penalty based on the target's ability to detect and attribute the source of a campaign. One can think of this penalty as some mixture of reputational, monetary, diplomatic, or other retributive sanctions imposed by the target state and any other states that back it. The government optimizes φ given the parameters.

As a stricter test of the theory, eliminate any agent drift θ so that non-state hackers are perfectly loyal to their government. Under the latter condition (coupled with the assumption that H's skill is equivalent to G's), default theory predicts that the government would strictly prefer to outsource. If we can identify any variation in φ at all, this suggests principal–agent problems are neither necessary nor sufficient conditions. As we will see, G's decision hinges on α.

For simplicity, assume the return on investment is linearly increasing in average attack scale ($\tilde{u} = 1$). Also, equalize the θ assumption by also assuming centralization costs are trivial for high-capacity governments ($\gamma\downarrow 0$), since these types have substantial resources at their disposal.[21] Like the model employed by Lindsay [13], this setup allows us to locate the optimal attack scale. Since I regard x as an exogenous parameter—i.e. the level of force required to achieve an effect—rather than a choice, it is safe to treat the probability of success, ψ, as a constant. Set $\psi = 1$ for the most interesting case. As a final simplification, assume technical attribution is certain whenever G attacks ($\tau = 1$), and that the penalty c is linearly increasing in scale as a function of the cumulative evidence left behind x. This implies a one-to-one increase in the target's ability to implicate the host state if the host state runs the campaign directly, since in this

[19] One could imagine that operational effectiveness varies by the level of centralization. For example, G's capabilities might be superior to H's. Then benefits would be modeled $\tilde{u}_\varphi x$. Because my objective is only to measure how attribution influences the decision to outsource, I hold constant skill equality between hackers and host government. There are theoretical reasons to justify this decision. In conventional domains, scholars have found a strong relationship between the level of host state support and the strength of proxy groups [93–98]. This could be because governments select the most qualified candidates, but also because support itself strengthens non-state actor capabilities. At low levels of support, namely tacit permissiveness, the relationship is probably weak or non-existent. If true for cyberspace, this would simply imply some mixed equilibrium.

[20] Assuming linearity in $f(\cdot)$.

[21] I presume a very small fixed cost $\gamma = \epsilon$ to eliminate the potential for mixed strategy equilibria. If the decisionmaker is ever indifferent between outsourcing/not, she will choose to outsource. This makes the test slightly more stringent.

Table 1: Parameter guide.

Term	Description	Range	Type
φ	Extent of outsourcing strategy	$\in [0, 1]$	G's strategy
ψ	Operational sophistication (probability of success)	$\in [0, 1]$	Exogenous
$\tilde{u}$	Return on scale of investment	$\in \mathbb{R}_{>0}$	Exogenous
x	Campaign investment	$\in \mathbb{R}_{\geq 0}$	Exogenous
θ	Distance between principal/agent ideal points	$\in \mathbb{R}_{\geq 0}$	Exogenous
γ	Fixed cost of centralization	$\in \mathbb{R}_{\geq 0}$	Exogenous
c	Penalty of being caught and attributed	$\in \mathbb{R}_{\geq 0}$	Exogenous
τ	Target's ability to attribute (technical attrib.)	$\in [0, 1]$	Exogenous
α	Target's willingness to accuse (political attrib.)	$\in \mathbb{R}_{\geq 0}$	Exogenous

case "responsibility" can be proven through technical means alone. Set x, $c = 1$.

This set of assumptions simplifies the equation to $U_G(\varphi) = x(2 - \varphi) - c(\alpha_v(\varphi) + x(1 - \varphi))$. In order to constitute an equilibrium where G outsources, the equation must satisfy the Incentive Compatibility Condition (ICC). This implies

$$\varphi > \frac{1}{a},$$

which illustrates the general relationship between outsourcing (φ) and the propensity of target states to react on the basis of circumstantial evidence. However, to make the analysis more interesting, relax the assumption that the scale of the campaign is endogenous to the amount of circumstantial evidence left behind. This instead gives us

$$U_G(\varphi) = x(2 - \varphi) - (\alpha_v(\varphi) + (1 - \varphi))\,,$$

where expected sanctions are the same for any level of operational investment. I study the relationship between political attribution and φ in Theorem 1.

Theorem 1. The higher the willingness of target states to hold the host state responsible for activity conducted within the latter's borders (α), the lower the host state's incentive to outsource (φ).

Proof of Theorem 1. Because this game is decision-theoretic with only one move, there is no Nash solution concept. The decision-maker locates the ideal level of φ via straightforward optimization. The proof is concise enough to show in the body of the paper.

With no loss of generality, square the cost term and divide by two to make the objective function continuously differentiable in the choice variable: $(\alpha_v(\varphi) + (1 - \varphi))^2$, which gives us the transformation

$$max\ U_G(\varphi) = x(2 - \varphi) - \tau\,(\alpha_v(\varphi) + (1 - \varphi))^2\,.$$

The first order condition (FOC) must isolate at least one extremum in order to find a critical point at which G will choose to switch between strategies. Differentiate the function with respect to φ:

$$\frac{\partial}{\partial\varphi}U_G(\cdot) = -(\alpha + 1)(\alpha\varphi + \varphi - 1) - x.$$

We can then solve for the optimal level of φ by simply setting the derivative equal to zero and rearranging algebraically for φ. This produces

$$\varphi^* = \frac{\alpha - x + 1}{(\alpha + 1)^2},\ \text{where } \alpha \geq 0 \text{ by assumption.} \tag{1}$$

The second order condition (SOC) confirms that φ^* is indeed a global maximum:

$$\frac{\partial^2}{\partial^2\varphi}U_G(\varphi) = -a^2 - 2a - 1,\ \text{which is } < 0. \tag{2}$$

We observe from Equation (2) that the slope is decreasing at U''_G, proving that φ^* maximizes G's payoff function. ■

Analysis

We wish to know how $U_G(\varphi^*)$ changes with α, x. Substituting φ^* into the original equation find calculate G's expected utility, we obtain

$$\begin{aligned} U_G(\varphi^*) = x\left[2 - \frac{\alpha - x + 1}{(\alpha + 1)^2}\right] \\ - x\left[\alpha\left(\frac{\alpha - x + 1}{(\alpha + 1)^2}\right) + \left(1 - \frac{\alpha - x + 1}{(\alpha + 1)^2}\right)\right], \end{aligned}$$

$$\text{which is equivalent to } U_G(\varphi^*) = \frac{\alpha x^2 + \alpha x + x}{\alpha^2 + 2\alpha + 1}. \tag{3}$$

Equation (1) models the optimal degree of outsourcing for any level of political blame or attack scale, assuming the best case scenario for all other parameters. Equation (3) models the utility G can expect to receive by adopting the optimal strategy. Although it is clear that outsourcing depends on the relationship between α and x, neither Equation (1) nor Equation (3) is immediately intuitive. The relationship becomes more apparent when α is φ^* is plotted. Figure 2 (right) shows the utility gained for different levels of φ. Outsourcing becomes less advantageous as α increases. In this model, the target state's α is common knowledge, so the host state can calibrate its outsourcing behavior accordingly. The plot on the left hand side of Fig. 2 maps the expected utility of optimal φ^*. As α increases, the expected gains of centralization ($\varphi^*\downarrow$) begin to exceed the gains from outsourcing ($\varphi^*\uparrow$). The only other way the host state can minimize costs is by reducing attack scale. Though the goal of this model is not to treat scale as a choice variable, some operations may require attacks of a scale that are cost-prohibitive given $\tau \times \alpha$, consistent with Lindsay [13]'s findings on deterrence. Imputing values can help give us more precise point estimates.

Example 1. Suppose the target state upholds a strict interpretation of the law on state responsibility, such that it is reluctant to assign blame on the basis of any non-technical evidence ($\alpha = 0$). Then the optimal level of outsourcing is $\varphi^* = x$ and G would receive $U_G(\varphi^*) = \tilde{u}x$ for any level of x. The value of x that maximizes this relationship is 1, so $\varphi^* = 1$. In other words, G gains plausible deniability for outsourcing and (absent any agent drift and assuming equal probability of success) the host stands to gain from outsourcing everything. It also implies (under identical conditions and assuming $\tilde{u} > 1$) that G would adopt a totally permissive attitude toward its hacker pool, since more activity yields a strictly better payoff. This outcome captures the conventional wisdom.

Example 2. How can we explain decisions not to outsource? Suppose a high willingness on the part of the target to politically attribute, e.g. $\alpha = 10$. Then $\varphi^* = (10x^2 + 11x)/121$, and the calculation becomes more complex.[22] Fixing $\varphi^* = 0.148$, $x^* = 0.895$ at their op-

[22] In this situation, G expects to receive the (rather hideous) payoff amount $\tilde{u}x((-0.0826x - 0.0909)x + 2) + x(x((-0.413x - 0.909)x + 0.409) +$

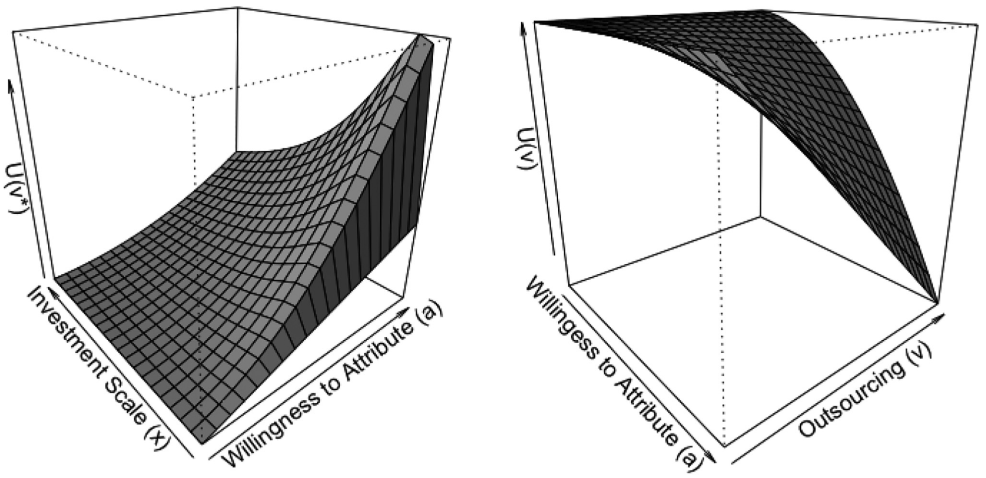

Figure 2: Total utility gained when behavior is optimal (left) and optimal level of outsourcing (right), given the victim's willingness to blame sponsors for their proxies. Notice utility is maximized when investment scale is minimized because lower investments make sponsor-proxy relationships easier to deny. Alone, this would suggest sponsors prefer more sophisticated proxies. In this model, however, the twist is that victims decide independently whether denials are plausible. When victims are quicker to blame sponsors, the cost of outsourcing approaches the cost of insourcing. In such a scenario, the risk of agent drift may serve as a tiebreaker. Attackers in this position are better off conducting their own in-house operations (whenever able).

tima (see previous footnote), G can expect to receive $0.148\tilde{u}$. Note that G's maximum payoff is strictly (and significantly) lower when $\alpha = 1$ no matter what its strategy. Payoffs diminish further as α increases or return on investment $\tilde{u}$ decreases, even when agents are perfectly loyal and highly capable, until rising penalties discourage the host government from outsourcing at all.

Discussion

In non-technical terms, the conditions under which it pays more to outsource than to centralize are narrowed by target states' increased willingness to assign political blame on the basis of technically insufficient or legally inadmissible evidence. Detection plus suspicions may sometimes be enough. The knowledge that one could be held accountable, whether publicly or privately, removes the incentives to outsource, especially when proxies are unsophisticated, careless, inefficient, lack fidelity to the cause, or have their own ulterior agenda. Even when proxies are perfect efficient and reliable, political attribution can dampen the attractiveness of outsourcing strategies.

In practice, the choice to outsource or centralize is not binary. States are free to pursue both strategies, with the understanding that resources must be divided. The model shows, however, that unless political attribution differs strongly between various operational applications, states are on average better off adopting the latter strategy. Empirical observations support these predictions, albeit only indirectly. The US Government is both increasingly willing to indict host states for their connection to proxies, and increasingly capable of providing technical evidence to substantiate its accusations. Similarly, punishments may be more regularized under the US' new policy of persistent engagement, and more states now see an interest in fostering robust, centrally operated cyber institutions [55] as opposed to decentralized strategies, such as outsourcing.

Pursuant to the specification used in Example 2, host states outsource only about 15% of the time. However, Lindsay [13] argues compellingly that a target's willingness to assign blame and/or retaliate could be positively related to the scale of the attack. If true, the results may even *underestimate* the influence of α, and therefore, overstate the rate of outsourcing. Operations that are serious enough to be of interest to the attacking state would likely invite in-kind retaliation even if the evidence were extremely flimsy. As technical attribution capabilities continue to improve and norms become more conducive as a consequence of emerging state practice, political attributions will only become more common.

Empirical Strategy

Researchers may not be able to test whether outsourcing is on the rise or decline directly, but they can ask about *insourcing*, since the latter is by definition more observable. We can also test whether known victim responses are associated with these changes. In this section, I provide evidence that they are. Limitations in the data notwithstanding, the findings lend support to the proposition that outsourcing may no longer pay because victims are increasingly willing to implicate states they suspect of being involved, even in the absence of smoking gun evidence. If true, it suggests a notable change is underway: if sponsors worry they will be held accountable regardless, then why not avoid the Promethean dilemmas associated with proxies?

There is no dataset that tracks negative variation in proxy relationships over time. As mentioned, Akoto [33] tracks onset of known proxy relationships only, not termination, and so cannot be used to test for a decline. I opt for the most up-to-date register of state-implicated cyber incidents, the CFR dataset (2005–2021). Incidents in the CFR data are ordered by the date news of an incident broke. I expand the data to include a row for every country (alleged sponsor) per day over this sixteen year period.

Each observation in CFR's dataset references two sources. I visited every source to code incidents by the level of evidence used to establish that a particular state was involved. In cases where the link did not specify, I investigated further. This generated an index ranging from 0 to 5, where 5 was direct evidence that particular government personnel were behind the attack. In cases where no evidence at all

1) − 0.5, which maximizes at $x \approx 0.895$. Suppose G has control over x, or at least has the power to increase or decrease x by permitting or restricting H from acting unilaterally. Substituting 0.895 into Equation (3), G maximizes utility by outsourcing at 0.148, or < 15% capacity.

for the relationship was provided, I coded this as a 0 (no evidence). See Table 2 for a breakdown of the coding scheme.[23] While this approach is far from perfect, the idea is that deniability is more objectively plausible at lower levels of the index (0–2), moderate when IoCs are good (3), and implausible when evidence directly implicates the state (4–5).

For information on victim responses, I rely on Hinck and Maurer [99]'s data on US indictments of foreign cyber actors. The first state-sponsored indictment was in 2014, 9 years after the first CFR-recorded cyber incident. I bring the Hinck and Maurer [99] dataset up to date (2022) by searching the DOJ's website for news and press releases under the category of cyber crime.[24]

Unsealed criminal indictments are a good test of the theory because they are necessarily public. Victims can respond in a variety of ways, from naming and shaming to launching a military counteroffensive [100]. Focusing on US DOJ indictments *per se* is a sensible approach. The high-profile nature of many of these indictments ensures that the attacking state received the message. The US in particular is also one of the biggest targets of state-sponsored cyberattacks, and has been especially active in indicting foreign suspects since 2014 [101]. Besides criminal indictments, the data also include information on whether suspects were extradited and tried, whether sanctions were levied, and, in rare known cases, whether the US hacked back. To code responses, each observation begins with a value of 1 by virtue of there having been a public acknowledgment. The value is increased by increments of 1 for each time the information indicates an additional step.[25] This generates a daily count variable ranging from 1 to 4.[26]

The data are then merged at the country-day level. Events are exceedingly rare given paucity of incidents relative to the total number of country-days between 2005 and 2021. This level of granularity is computationally challenging to analyze. I therefore consolidate the data at an annual level with no expected loss in generality.

In Fig. 3, we can see how insourcing has changed over time. Lighter colors indicate more direct evidence that state personnel, rather than non-state proxies, were involved (index values: 4–5). The second category, technical indicators, encompasses technical indicators like network traces and IoCs (index values: 2–3). These technical indicators can be contrasted to the third category, unconfirmed suspicions (index values: 0–1). Incidents with index values of 4 and especially 5 are of particular interest because, as opposed to proxy operations, government involvement is categorically undeniable no matter how strict the burden of proof is set. While a greater share of technical evidence has been offered since 2013, direct evidence has become by far the most common form of proof.[27] Footnoted caveats notwithstanding, this should imply that states are conducting more operations with their own uniformed personnel. This relationship holds for each of the "big four" US adversaries.

Analysis

To explain the patterns in Fig. 3, we want empirical models that can provide answers to two questions. First, do US responses contribute to more outsurcing (as sponsors seek more plausible deniability) or more insourcing (because deniability is not actually plausible)? Second, just how much deniability do proxies afford, anyway? In other words, does the US only respond when it can present direct evidence that a sponsoring state was involved? I estimate two pairs of lagged fixed effects models designed to address these questions.

In the first pair, I test whether US responses are associated with a change in future adversary behavior. In the second pair, I test the reverse: whether evidence of direct state involvement is any more likely to provoke a subsequent US response. Together, these tests tell a compelling story. If the US is just as likely to exact punishment regardless of whether its adversaries rely on proxies, then the use of proxies for plausible deniability no longer pays, and adversaries should insource more going forward. This is consistent with the findings.

For tractability, I make a necessary assumption. As discussed, it is not possible to measure outsourcing *per se* because the more plausibly deniable operations are, the more likely they are to drop out of our datasets by pooling with non-state sponsored operations. I assume that aggressors have a consistently high demand for cyber operations in each period, which they can achieve either by insourcing or outsourcing (country/year fixed effects also help account for this unobserved variation, as discussed within). We cannot study outsourcing directly, but as long as this assumption holds, *more insourcing* should imply *less outsourcing*.

Next, in order to construct the dependent variable for the first set of models (and independent variable in the second set of models), I dichotomize the evidence index. Evidentiary index values between 4 and 5 are coded as proof that country i was implicated in year j. In order to give plausible deniability maximal benefit of the doubt, values below this threshold, including strong IoCs, are coded as no proof having been observed. This distinguishes between direct involvement (undeniable) and indirect involvement (potentially plausibly deniable). It also ensures we are measuring insourcing *per se*, and not simply an accumulation of indirect, lower-level evidence. Without a clear way to quantify the scale or severity of a response, the other variable, US response, is also dichotomized for each country-year.

I include 1-year lags for each variable, Public Response and Observe Insourcing, when used as predictors. This ensures our estimation of the direction of the relationship in each model is one-way and sequential. It allows adversaries time to learn and adjust after

[23] The rest of the scheme is as follows: (1) *cui bono* was specifically invoked, (2) network traces, (3) more sophisticated IoCs were specifically discussed, and (4) a particular government organ or known, regular contractor was implicated.

[24] https://www.justice.gov/news

[25] These include whether the source of the attack/incident was taken down; whether sanctions were levied; whether the suspects were placed on the "Most Wanted" list or a reward was offered for their capture; and whether they were extradited, arrested, and/or sentenced.

[26] It stands to reason that, if anything, the Hinck and Maurer [99] data might understate the intensity of the US response. One can imagine that many if not most US Government hackbacks are undisclosed to the public.

[27] Two important caveats should be discussed. First, it is possible that state personnel were always this involved, that the US has always known this, and it is simply more willing to disclose it since 2014. This is a shortcoming that is impossible to address because the political processes contributing to disclosures are unobserved. If true, the observed increase in evidence since 2014 is a function of victim confidence rather than actual changes in activity. However, given that 100% of attributed incidents were insourced in the last year of the dataset, insourcing as a proportion at least cannot have declined. A second possibility, that attribution technology itself has improved, so the US, even if always in theory willing, is now more *able* to attribute. This is less of a concern. Attribution has certainly improved over time, but that should mostly inflate the number of observed 3s (IoCs) as lower levels of evidence (0–2) become provable. Here, we are principally concerned with whether the US is more likely to observe and respond to direct (≥ 4) or indirect evidence (< 4).

Table 2: Quality of evidence cited in attribution for publicly known incidents in CFR dataset, indexed on a {0:5} scale. Higher values indicate more obvious sponsor involvement. The last column indicates the theorized level of political cover a sponsor obtains for the operation *vis-a-vis* a suspicious target state. **Bolded values** are dichotomized for Models 1 and 2 to differentiate physical evidence (about a person/agency behind the computer) from technical evidence (about the computer system alone).

Evidence	Value		Suspicious target
Uniformed personnel	(5)	**Direct proof**	**Implausibly deniable**
Agency/contractor	(4)	**Strong proof**	**Implausibly deniable**
IoCs: known group	(3)	Technical proof	Questionably deniable
IoCs: location	(2)	Technical proof	Plausibly deniable
Cui bono	(1)	Circumstantial proof	Plausibly deniable
No evidence	(0)	No proof	Plausibly deniable

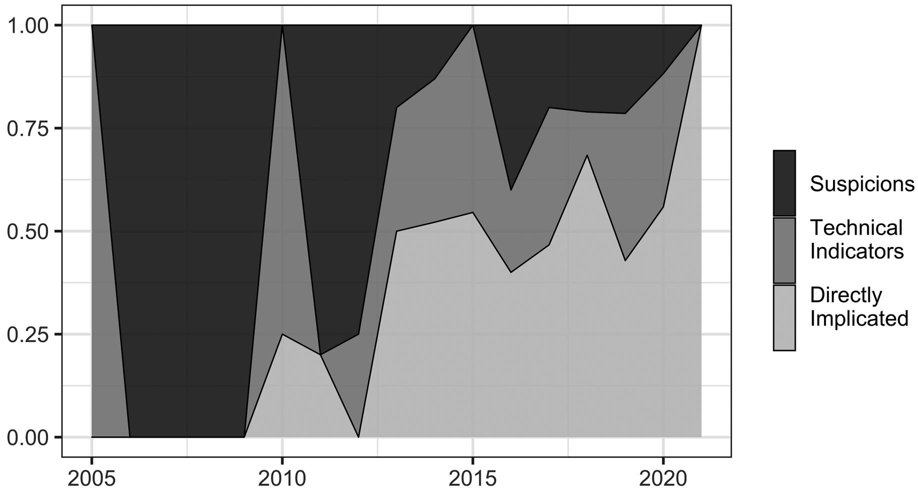

Figure 3: Type of evidence used in support of attributions. The density plot is filled to control for variation in the number of detected operations per year. Lighter shades indicate stronger evidence that a government was involved. Evidence that directly implicates a government culprit (i.e. undeniable involvement) is increasing in share over time.

Table 3: Linear fixed effects models. Results from four specifications with cluster-robust standard errors reported. Estimands of interest are highlighted. The first two models estimate the effect of US responses on other countries' propensity to insource operations in the next period. The latter two examine how direct evidence about who an attacker was in that period contributes to the US Government's willingness to respond publicly in the following period. Even columns present country fixed effects with a time control. Two-way fixed effects are presented in the even columns.

	Dependent variables:			
	Observe insourcing		Public response	
	Sponsor FE (1)	Two-way FE (2)	Sponsor FE (3)	Two-way FE (4)
Public response ($t-1$)	0.049*** (0.016)	0.050*** (0.017)		
Obs. insourcing ($t-1$)			0.028 (0.043)	0.035 (0.042)
Year (*Control*)	0.015*** (0.005)		0.013** (0.006)	
Incident (*Control*)	0.114*** (0.015)	0.112*** (0.010)	0.088*** (0.022)	0.082*** (0.021)
Observations	186	186	186	186
R^2	0.450	0.304	0.355	0.207
Adjusted R^2	0.409	0.180	0.307	0.065

Note*:* *$P < 0.1$; **$P < 0.05$; ***$P < 0.01$.

the US responds (Models 1 and 2) and time enough for the US to detect and marshal a public response to adversary intrusions (Models 3 and 4). Fixed effects are also used to control for any unobserved variation in the predictors, i.e. specific to a sponsoring country or year. Controls are included for the number of incidents that occurred and the year in which they occurred (in the one-way fixed effects models). Without a theoretical basis for adding further controls, I opt for parsimony.

Table 3 presents the results from all four models. The base models (even columns) use one-way (sponsoring country) fixed effects with a time control. Two-way (country/year) fixed effects are presented in the odd columns as a robustness check.[28] Consistent with best practices in the literature, robust standard errors, clustered at the country level, are reported. The first two models estimate the effect of US responses on other countries' propensity to insource operations in the next period. The latter two examine how direct evidence about who an attacker was in that period contributes to the US Government's willingness to respond publicly in the following period. The estimands in all four models are highlighted.

First examine Model 1, the effect of US responses (with a one-year lag) on insourcing. The dependent variable, Observe Insourcing, is whether publicly disclosed incidents in a given year were known to have been conducted directly by another state's own agencies or uniformed personnel. Model 1 shows that US responses in past periods ($t - 1$) are positively associated with decisions by other countries to insource in subsequent periods ($P < 0.01$).[29] Since the dependent variable ranges between 0 and 1, one way to interpret the coefficients is that public responses are associated with a 5% increased likelihood that an adversary will instead rely on its own operatives in future periods.[30] Model 2 shows that this finding is robust to two-way fixed effects. Returning to the assumption, if we believe that the demand for cyber operations is inelastic, then an increase in insourcing would imply a decrease in outsourcing.

Models 3 and 4 test the second question: at what evidentiary threshold is the US willing to cast blame? The independent and dependent variables (Observe Insourcing and Public Response, respectively) are on the same scale, but this time Insourcing is lagged by 1 year. We see that there is no significant association between the US having obtained irrefutable evidence of a sponsor's involvement and its willingness to respond publicly ($P = 0.35$). The US may be able to more accurately detect and attribute systems, but its ability to implicate individuals is apparently unrelated how eagerly it retaliates against the usual suspects.

Why might states outsource, if not plausible deniability? There are other potential reasons why states might decide to insource more in later years, but these would not upset the findings. For example, target hardening or choice of target might necessitate more sophisticated operations. Hacker collectives in certain countries may no longer be reliably sympathetic. States may want to use new capabilities they have gained. But even if there are additional reasons why states insource, by doing so, plausible deniability is compromised.

Moreover, the inclusion of fixed effects (in Models 1 and 2) should increase our confidence that the US response has strongly contributed to this trend. This approach ensures that any unobserved factors unique to a sponsor or time period are isolated and eliminated when they might affect the outcome. The same is true for Models 3 and 4, which estimate how the US responds to different countries over time depending on the evidence. Because other possible factors are accounted for by fixed effects, it "reduces concerns that omitted variables drive any associations between dependent and independent variable" [104].

Though caution should be used in interpreting these results given limitations in the data and cyber secrecy more generally, they offer tentative support for the theory that proxies may no longer pay. The US is far from the only cyber victim, but it is among the most vocal [105]. US officials have sometimes insisted that 'no one action can change [adversary] behavior in cyberspace' [106]. Yet the US now seems to be realizing that complicity is in the eye of the beholder. US adversaries surely perceive this change, as they too have recently begun to publicly attribute [107,108]. If interest in completing operational objectives through cyber means has been consistent since 2014, more insourcing would imply less outsourcing. This is unfortunately untestable. However, as long as this condition holds, the empirical results would seem to confirm that sponsors are adjusting their behavior by outsourcing less and insourcing more.

Conclusion

Scholars have long been interested in the conditions under which new cyber norms can emerge and proliferate [109–112]. Naval privateering was once common; states used letters of marque to commission pirates and mercenaries to conduct raids on their adversaries, but this strategy became obsolete in part because victims changed the norm by holding state sponsors responsible [113,114]. Could new norms around state responsibility for cyberattacks also be coalescing around lower evidentiary barriers? This paper cannot, and does not, definitively contend that outsourcing is *in fact* on the decline. Instead, it is intended to caution against overreliance on untested assumptions about what states are doing and why they do it.

On the open source level, it is impossible to say with certainty whether outsourcing is increasing, decreasing, or staying the same. Outsourcing could be declining because deniability no longer plausible. But if proxies really do still convey deniability, and this deniability is plausible, then the patterns in Fig. 2 make sense: operations would be impossible to connect to a sponsoring government, so observers would (mistakenly) perceive no increase in outsourcing trends. In short, observed patterns are produced by one of two mutually exclusive data-generating processes: a *true* decline, or a turn away from *ineffective* proxies in favor of even stealthier alternatives.

This potential for observational equivalence is all the more reason why we should be cautious before uncritically adopting the "plausible deniability" assumption. Embracing multiple possibilities can help us understand the sensitivity of our theories to our underlying assumptions. Moreover, the statistical model presents at least some empirical support for the idea that the plausible deniability conjecture is misplaced, and that outsourcing might no longer pay. Trying to measure outsourcing directly is problematic for reasons dis-

[28] Recent research has strongly recommended one-way fixed effects models over two-way specifications [102]. Because two-way specifications remain popular in political science, I include both in Table 3 to demonstrate robustness. Although a logistic model could also be appropriate for binary outcomes, OLS is preferable in this case. It is now understood that OLS is the best unbiased estimator (linear or otherwise) [103]. Linear coefficients are also more easily interpreted.

[29] The reliability of this estimate assumes that the US was equally likely to acknowledge any evidence it had of *direct* involvement across time. This is an unavoidable shortcoming in reporting-based data given the level of secrecy surrounding internal public attribution decisions. On the upside, because the dependent variable was dichotomized before being summed and normalized by year, it does not assume that the US was any more or less likely to disclose *indirect* or piecemeal evidence, such as IoCs. The first assumption is probably less tenuous than the latter.

[30] Caution should be used in this interpretation, since fixed effects limit the number of comparisons. The difficulty in making substantive comparisons between units is a principal downside of fixed effects [104]. The need to compensate for the strong possibility of selection bias outweighs limitations in interpretation.

cussed above. A more tractable approach is to measure the relative propensity of attackers to conduct their *own* operations. The more involved the sponsoring government is, the more likely the operation is to show up in data on state-linked attacks. This is because the leap between attack detection and political attribution is easier to make. Where conduct is known to be direct, involvement in detected attacks is objectively undeniable. If proxies are at best substitutable for government operations, but governments are observed to be doing more on their own, then this could imply a decline in outsourcing.

As opposed to data on outsourcing, the data on insourcing—which we *can* test—do seem to tell a compelling story. Attackers are conducting a greater proportion of attacks themselves. We can also test whether this change has anything to do with implausible deniability. The results indicate that public responses by powerful victims like the US are statistically predictive of more direct operational involvement by government operatives in subsequent attacks. Moreover, since 2014, the quality of evidence regarding state involvement seems to have no predictive value in whether victims respond. Attribution may be getting more sophisticated, but victims' suspicions are basis enough. In either case, attackers lose their political cover from outsourcing.[31]

Without plausible deniability, would-be sponsors face a variety of disincentives. In conventional spaces, some speculate that outsourcing generally leads to wider disruption and collateral damage [20], whereas state control minimizes these byproducts [115]. In the cyber context, host states may not wish to risk bringing neutral parties into the fray, accidentally damaging assets in blue space, or violating emerging norms of international humanitarian law, such as the prohibition on attacking critical infrastructure in peacetime. Mounting evidence has also led some researchers and policymakers to cautiously conclude that cyber conflict is not inherently escalatory, anyway [56,58]. If true, why jeopardize operational efficiency by outsourcing to outside actors who might be less-than-reliable? In converse, command centralization may come with advantages in terms of the division of skilled labor, economies of scale, and operational coherence.

If not plausible deniability, then why outsource? It is possible that outsourcing is not always designed to affect a target but rather to rally nationalism for internal cohesion [72,116,117] or external signaling purposes [26,118]. And at times and in places where online nationalism is elevated (see e.g. [25]), host states should find it easier than ever to co-opt sympathetic hackers. When non-state proxies are ideologically aligned with the state's objectives [93], or when the host state lacks capacity to conduct covert operations of its own [20,119], a state can preserve its own resources by outsourcing to them. In some rare cases, proxies might even be more sophisticated than the state and therefore able to accomplish things the state cannot. State support can also transform a group's goals [120], helping not only to minimize agency drift but to stem it (cf. [8]). However, if states are using proxies for reasons *other* than plausible deniability, then this would be reflected in the data by an increasing trend. This is not what the data in Fig. 2 depict.

Formalizations such as the decision-theoretic model used in this paper offer several advantages to the study of cyber conflict in areas where measurement, selection, and reporting bias are presumed to be especially severe. Herr et al. [121], Gorwa and Smeets [122], and others have issued a call to action to cyber researchers to pay greater attention to the internal validity of their theories. Theories should be expected to generate clean and testable empirical predictions if the science of cyber conflict is to advance [122]. Formal models like the one employed in this paper, while underutilized, are uniquely suited for studying relationships in which data are unreliable or elusive. Formalization allows researchers to demonstrate exactly why some explanations can be safely ruled out and others cannot. They can also help isolate a set of valid hypotheses, assess the relative importance of certain variables, and generate falsifiable predictions for when better data do become available.

This research also contributes to longstanding debates on the role of the state in cyberspace and its relationship to private actors [123,124]. In line with previous observations about how cyber research and policy can sometimes be led astray when basic assumptions are not questioned [125], cyber conflict scholarship is on the cusp of a more empirical turn. In addition to the rich body of qualitative work already prevalent in cyber conflict studies, quantitative and large-*n* studies can provide special insight into broader patterns (e.g. [54,56]). Moving forward, tools that ensure that the theoretical predictions underpinning empirical studies are clear, parsimonious, and internally consistent will be especially valuable, regardless of methodological orientation.

Acknowledgements

The author is indebted to William Akoto, Erik Gartzke, Jason Healey, Nadiya Kostyuk, Jon Lindsay, Mike Poznansky, Jack Snyder, Brandon Valeriano, Chris Whyte, Josephine Wolff, JD Work, participants of the Digital Issues Discussion Group (DIDG), and four anonymous reviewers for helpful comments that improved this manuscript.

References

1. Mumford A. Proxy warfare and the future of conflict. *RUSI J* 2013; **158**: 40–6. doi: 10.1080/03071847.2013.787733.
2. Borghard E, Friends with benefits? Power and influence in proxy warfare. Ph.D. Thesis, Columbia University, 2014.
3. Salehyan I, Siroky D, Wood RM. External rebel sponsorship and civilian abuse: a principal-agent analysis of wartime atrocities. *Int Org* 2014; **68**: 633–61.
4. Szekely O. A friend in need: the impact of the Syrian Civil War on Syria's Clients (a principal—agent approach). *Foreign Pol Anal* 2016; **12**: 450–68.
5. Abbott KW. Economic sanctions and international terrorism. *Vand J Transnat'l L* 1987; **20**; 289.
6. Kirchner M. "A good investment?" state sponsorship of terrorism as an instrument of Iraqi foreign policy (1979–1991). *Camb Rev Int Aff* 2014; **27**: 521–37. doi: 10.1080/09557571.2013.839629.
7. Byman D, Kreps SE. Agents of destruction? Applying principal-agent analysis to state-sponsored terrorism. *Int Stud Perspect* 2010; **11**: 1–18.
8. Borghard ED, Lonergan SW. Can states calculate the risks of using cyber proxies?. *Orbis* 2016; **60**: 395–416.
9. Applegate S. Cybermilitias and political hackers: use of irregular forces in cyberwarfare. *IEEE Secur Priv* 2011; **9**: 16–22.
10. Poznansky M, Perkoski E. Rethinking secrecy in cyberspace: the politics of voluntary attribution. *J Glob Secur Stud* 2018; **3**: 402–16.
11. Brown JM, Fazal TM. #SorryNotSorry: why states neither confirm nor deny responsibility for cyber operations. *Eur J Int Secur* 2021; **6**: 1–17.
12. Baliga S, Bueno de Mesquita E, Wolitzky A. Deterrence with imperfect attribution. *Am Polit Sci Rev* 2020; **114**: 1155–78.

[31] It is conceivable that outsourcing and insourcing are *both* increasing, and that outsourcing to plausibly deniable actors is increasing at a faster clip. This would undermine the results. Cyber operations have been consistently attractive to governments over the past decade. Thus, while possible, this alternative seems unlikely.

13. Lindsay JR. Tipping the scales: the attribution problem and the feasibility of deterrence against cyberattack. *J Cybersecur* 2015; **1**: 53–67.
14. Cormac R, Aldrich RJ. Grey is the new black: covert action and implausible deniability. *Int Aff* 2018; **94**: 477–94.
15. Lin-Greenberg E, Milonopoulos T. Private eyes in the sky: emerging technology and the political consequences of eroding government secrecy. *J Conf Resol* 2021; **65**: 1067–97.
16. Vaynman J. Better monitoring and better spying: the implications of emerging technology for arms control. *Texas Natl Secur Rev* 2021; **4**: 33–56.
17. Pitrelli MB. 'For the first time in history anyone can join a war': volunteers join Russia-Ukraine cyber fight. CNBC. Section: Technology. 2022. CNBC. https://www.cnbc.com/2022/03/14/volunteers-sign-up-to-help-in-cyberwars-between-russia-and-ukraine-.html. (11 March 2022, date last accessed).
18. Fedorov M. We are creating an IT army. We need digital talents. 2022. All operational tasks will be given here: https://t.me/itarmyofurraine. There will be tasks for everyone. We continue to fight on the cyber front. The first task is on the channel for cyber specialists. @FedorovMykhailo. https://twitter.com/FedorovMykhailo/status/1497642156076511233. (11 March 2022, date last accessed).
19. Stephenson E. Does United Nations war prevention encourage state-sponsorship of international terrorism - an economic analysis. *Virg J Int Law* 2003; **44**: 1197.
20. Salehyan I, Gleditsch KS, Cunningham DE. Explaining external support for insurgent groups. *Int Org* 2011; **65**: 709–44.
21. Hoffman B. *Inside Terrorism. REV - revised, 2*. New York: Columbia University Press, 2006.
22. Findley MG, Piazza JA, Young JK. Games rivals play: terrorism in international rivalries. *J Polit* 2012; **74**: 235–48.
23. Conrad J. Interstate rivalry and terrorism: an unprobed link. *J Conf Resolut* 2011; **55**: 529–55.
24. Singer PW, Friedman A. *Cybersecurity and Cyberwar: What Everyone Needs to Know®*. Oxford: Oxford University Press, 2013.
25. Gries PH. *China's New Nationalism*. 1st edn. Berkeley, Los Angeles, London: University of California Press; 2005, 226.
26. Weiss JC. *Powerful Patriots: Nationalist Protest in China's Foreign Relations*. 1st edn. New York: Oxford University Press, 2014, 360.
27. Lemos R. Defacements rise in China hacker war. CNET. library catalog. www.cnet.com. https://www.cnet.com/news/defacements-rise-in-china-hacker-war/. (11 March 2022, date last accessed).
28. Mulvenon JC, Tanner MS, Chase MS. *et al.*, Option four: Chinese network-centric warfare. In: *Chinese Responses to U.S. Military Transformation and Implications for the Department of Defense*, 1 edn. Santa Monica: RAND Corporation, 2006. 133–44.
29. Bahovski E. Francis Maude: the cyber-attack against Estonia was a big wake-up call for the world. Diplomaatia. Library Catalog: icds.ee. Tallinn: International Centre for Defence and Security (ICDS). 2012.
30. Shachtman N. Kremlin Kids: we launched the Estonian cyber war. Boone: WIRED, 2009.
31. Nye JS. Cyber power. Cambridge: Belfer Center for Science and International Affairs, Harvard Kennedy School, 2010.
32. Egloff FJ. *Semi-State Actors in Cybersecurity*. Oxford: Oxford University Press, 2021. 294.
33. Akoto W. Accountability and cyber conflict: examining institutional constraints on the use of cyber proxies. In: *Conflict Management and Peace Science*. Thousand Oaks: SAGE Publications Ltd, 2021. 07388942211051264.
34. Healey J. The spectrum of national responsibility for cyberattacks. *Brown J World Affairs* 2011; **18**: 57–70.
35. Maurer T. *Cyber Mercenaries: The State, Hackers, and Power*. Cambridge, New York, Port Melbourne, New Delhi, Singapore: Cambridge University Press, 2018. 268.
36. Damrosch L, Murphy S. International Law. 6th edn. St. Paul: West Academic Publishing, 2014. 1250.
37. Maurer T. Cyber proxies and their implications for liberal democracies. *Wash Q* 2018; **41**: 171–88. doi: 10.1080/0163660X.2018.1485332.
38. Schmoldt J. The rising power of cyber proxies, vol. Thaddeus Eze (ed.). In: *Proceedings of Twentieth European Conference on Cyber Warfare and Security*, Chester. 2021. p. 646.
39. Atwell K, Portzer JM, McCurdy D. Negotiating [Im]plausible deniability: strategic guidelines for U.S. engagement in modern indirect warfare. *PRISM* 2021; **9**: 112–21.
40. Feuer SV. From the shadows to the front page: state use of proxies for cyber operations. Stanford: Freeman Spogli Institute for International Studies, Stanford University, 2020.
41. Collier J. Proxy actors in the cyber domain: implications for state strategy. *St Antony's Int Rev* 2017; **13**: 25–47.
42. Maurer T. *Cyber Proxies and the Crisis in Ukraine, vol. Cyber war in perspective: Russian aggression against Ukraine, Kenneth Geers (ed.). OCLC: 960393919*. Tallinn: NATO Cooperative Cyber Defence Centre of Excellence, 2015.
43. Carlin JP. Detect, disrupt, deter: a whole-of-government approach to national security cyber threats. *Harv Natl Secur J* 2015; **7**: 391–436.
44. Denning D. Cyber conflict as an emergent social phenomenon. In: Holt TJ, Hlubik Schell B (eds.), *Corporate Hacking and Technology-driven Crime*. Pennsylvania: IGI Global, 2011.
45. Segal A. Beware the patriotic geek: the risk of cyber militias in Asia. New York: Council on Foreign Relations. 2012. Library Catalog: www.cfr.org. https://www.cfr.org/blog/beware-patriotic-geek-risk-cyber-militias-asia URL. (11 March 2022, date last accessed).
46. Klimburg A. Mobilising cyber power. *Survival* 2011; **53**: 41–60.
47. Trozzo E. *The Cyberdimension: A Political Theology of Cyberspace and Cybersecurity*. Eugene: Wipf and Stock Publishers, 2019. p. 293.
48. Ottis R. From pitch forks to laptops: volunteers in cyber conflicts. In: *Proceedings of the Conference on Cyber Conflict*, Tallinn. 2010. p. 97–109.
49. Pagliery J. Meet the vigilante who's hacked jihadist websites for years; CNNMoney. Library Catalog: money.cnn.com. 2015. https://money.cnn.com/2015/01/16/technology/security/jester-hacker-vigilante/index.html. (11 March 2022, date last accessed).
50. Thornburgh N. The invasion of the Chinese cyberspies. Time, 2005.
51. Winter J. Patriot hacker 'The Raptor' gains flock of followers after FoxNews.com report. Fox News. 2015. Last Modified: 2015-03-26T17:42:51-04:00 Library Catalog: www.foxnews.com. https://www.foxnews.com/us/patriot-hacker-the-raptor-gains-flock-of-followers-after-foxnews-com-report. (11 March 2022, date last accessed).
52. Bob YJ, Joffre T. Iran's Mahan Air hit by cyberattack, materials allegedly linked to IRGC. The Jerusalem Post | JPost.com. 2021.
53. NIPC encourages heightened cyber security as Iraq - US tensions increase. National Infrastructure Protection Center (NIPC) Advisory 03-002. National Infrastructure Protection Center (NIPC), 2003. https://www.2600.com/news/mirrors/ www.nipc.gov/warnings/advisories/2003/03-002.htm. (11 March 2022, date last accessed).
54. Kostyuk N, Zhukov YM. Invisible digital front: can cyber attacks shape battlefield events? *J Conf Resol* 2019; **63**: 317–47.
55. Kostyuk N. Deterrence in the cyber realm: public versus private cyber capacity. *Int Stud Quart* 2021; **65**: 1151–62.
56. Valeriano B. Cyber war versus cyber realities: cyber conflict in the international system. 1st edn. Oxford, New York: Oxford University Press, 2015. p. 288.
57. Council on Foreign Relations. Tracking state-sponsored cyberattacks around the world. 2021. https://www.cfr.org/cyber-operations. (11 March 2022, date last accessed).
58. Valeriano B, Jensen B, Maness RC. Cyber strategy: the evolving character of power and coercion. New York: Oxford University Press, 2018. p. 320.

59. Kramer F, Starr SH, Wentz L, eds, *Cyberpower and National Security*. 1st edn. Washington: Potomac Books, 2009.
60. Libicki MC. Cyberdeterrence and cyberwar. Santa Monica: Rand Corporation, 2009. p. 239.
61. Betz D, Stevens T. *Techniques for Cyber Attack Attribution*. Alexandria: Institute for Defense Analyses, 2003. p. 82.
62. Lin H. Attribution of malicious cyber incidents: from soup to nuts. *J Int Aff* 2016; **70**: 75–137.
63. Rid T, Buchanan B. Attributing cyber attacks. *J Strat Stud* 2015; **38**: 4–37. doi: 10.1080/01402390.2014.977382.
64. Canfil JK, *Intelligence and Adversaries: What Do We Know?*. New York: Cyber Conflict Studies Association (CCSA), 2016.
65. Carnegie A, Carson A. The disclosure dilemma: nuclear intelligence and international organizations. *Am J Polit Sci* 2019; **63**: 269–85. https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12426. (11 March 2022, date last accessed).
66. Murphy SD. *Principles of International Law*. Toronto: Thomson/West, 2012. p. 575.
67. Canfil JK. A framework for assessing foreign state complicity: a framework for assessing foreign State complicity. *J Int Aff* 2016; **70**: 217–26.
68. Weber V. *States and Their Proxies in Cyber Operations*. Washington: Lawfare, 2018.
69. Cole A, Healey J. United States should discourage “patriotic hackers” from attacking North Korea. Atlantic Council. 2014. Library Catalog: www.atlanticcouncil.org. https://www.atlanticcouncil.org/blogs/new-atlanticist/us-should-discourage-patriotic-hackers-from-attacking-north-korea/. (11 March 2022, date last accessed).
70. Segal A. *The Hacked World Order: How Nations Fight, Trade, Maneuver, and Manipulate in the Digital Age*. 1 edn. New York: PublicAffairs, 2016.
71. Wu X. *Chinese Cyber Nationalism: Evolution, Characteristics, and Implications: Evolution, Characteristics, and Implications*. Washington: Lexington Books, 2007. p. 280.
72. Hang R. Freedom for Authoritarianism: Patriotic Hackers and Chinese Nationalism. The Yale Review of International Studies. Library Catalog: yris.yira.org. 2014. http://yris.yira.org/essays/1447. (11 March 2022, date last accessed).
73. Henderson S. *The Dark Visitor*. 1st edn. New York: Scott Henderson, 2007. p. 148.
74. Fletcher O. Patriotic Chinese hacking group reboots. WSJ. ISSN: 0099-9660 Library Catalog: blogs.wsj.com Section: World, 2011. https://blogs.wsj.com/chinarealtime/2011/10/05/patriotic-chinese-hacking-group-reboots/. (11 March 2022, date last accessed).
75. Newman LH. China escalates hacks against the US as trade tensions rise. San Francisco: WIRED. 2018.
76. Egloff FJ, Smeets M. Publicly attributing cyber attacks: a framework. *J Strat Stud* 2021; 1–32. doi: 10.1080/01402390.2021.1895117.
77. Egloff FJ. Public attribution of cyber intrusions. *J Cybersecur* 2020; **6**: tyaa012.
78. Department of Justice Office of Public Affairs. U.S. charges five Chinese military hackers for cyber espionage against U.S. corporations and a labor organization for commercial advantage. Washington. 2014.
79. Berghel H. The equifax hack revisited and repurposed. *Computer* 2020; **53**: 85–90.
80. Department of Justice Office of Public Affairs. DOJ press release: U.S. charges three Chinese hackers who work at internet security firm for hacking three corporations for commercial advantage. Washington. 2017.
81. Department of Justice Office of Public Affairs. DOJ press release: two Chinese hackers associated with the Ministry of State Security charged with global computer intrusion campaigns targeting intellectual property and confidential business information. Washington. 2018.
82. Department of Justice Office of Public Affairs. DOJ press release: member of sophisticated China-based hacking group indicted for series of computer intrusions, including 2015 data breach of Health Insurer Anthem Inc. affecting over 78 million people. Washington. 2019.
83. Department of Justice Office of Public Affairs. DOJ press release: attorney general William P. Barr announces indictment of four members of China’s military for hacking into equifax. Washington. 2020.
84. Nakashima E, Lynch DJ. U. S. charges Chinese hackers in alleged theft of vast trove of confidential data in 12 countries. Washington Post. 2018. https://www.washingtonpost.com/world/national-security/us-and-more-than-a-dozen-allies-to-condemn-china-for-economic-espionage/2018/12/20/cdfd0338-0455-11e9-b5df-5d3874f1ac36_story.html. (11 March 2022, date last accessed).
85. Harknett R, Fischerkeller M. Persistent engagement and tacit bargaining: a path toward constructing norms in cyberspace. Lawfare. 2018. https://www.lawfareblog.com/persistent-engagement-and-tacit-bargaining-path-toward-constructing-norms-cyberspace. (11 March 2022, date last accessed).
86. Miller JN, Pollard NA. Persistent engagement, agreed competition and deterrence in cyberspace. Lawfare. 2019. https://www.lawfareblog.com/persistent-engagement-agreed-competition-and-deterrence-cyberspace.
87. Waxman MC. Cyber-attacks and the use of force: back to the future of Article 2(4). *Yale J Int Law* 2011; **36**: 421–60.
88. Kaminska M. Restraint under conditions of uncertainty: why the United States tolerates cyberattacks. *J Cybersecur* 2021; **7**: tyab008.
89. Egloff FJ, Dunn Cavelty M. Attribution and knowledge creation assemblages in cybersecurity politics. *J Cybersecur* 2021; **7**: tyab002.
90. Axelrod R, Iliev R. Timing of cyber conflict. *Proc Natl Acad Sci* 2014; **111**: 1298–303.
91. Merlevede J. Benjamin J. Jens G. Tom H. Exponential Discounting in Security Games of Timing. *Journal of Cybersecurity* 2021; 7(1): tyaa008. https://doi.org/10.1093/cybsec/tyaa008.
92. Gailmard S, Patty JW. Preventing prevention. *Am J Polit Sci* 2019; **63**: 342–52.
93. Byman D, Scheuer M, Lieven A, Lang WP. Iraq, Afghanistan and the war on “Terror”. *Mid East Pol* 2005; **12**: 1–24.
94. Quillen C. A historical analysis of mass casualty bombers. *Stud Conflict Terror* 2002; **25**: 279–92.
95. Hoffman B. Terrorism and weapons of mass destruction: an analysis of trends and motivations. Santa Monica: RAND Corporation, 1999.
96. Benjamin D, Simon S. *The Age of Sacred Terror: Radical Islam’s War Against America*. Reprint edn. New York: Random House Trade Paperbacks, 2003. 560.
97. Merari A. Terrorism as a strategy of insurgency. *Terror Polit Violence* 1993; **5**: 213–51. doi: 10.1080/09546559308427227.
98. Anderson SK. US counterinsurgency vs Iranian-sponsored terrorism. *Low Inten Conflict Law Enforc* 2002; **11**: 254–70. doi: 10.1080/0966284042000279027.
99. Hinck G, Maurer T. Persistent enforcement: criminal charges as a response to nation-state malicious cyber activity. *J Natl Secur Law Pol* 2019; **10**: 525–62.
100. Gomez MA, Whyte C. Unpacking strategic behavior in cyberspace: a schema-driven approach. *J Cybersecur* 2022; **8**: 1–16.
101. Franceschi-Bicchierai L. Chinese cybersecurity company doxes apparent NSA hacking operation. Vice. 2022. https://www.vice.com/en/article/v7dxg3/chinese-cybersecurity-company-doxes-apparent-nsa-hacking-operation. (11 March 2022, date last accessed).
102. Kropko J, Kubinec R. Interpretation and identification of within-unit and cross-sectional variation in panel data models. *PLOS ONE* 2020; **15**: e0231349.
103. Hanson BE. A Modern Gauss-Markov Theorem. *Econometrica* 2021. (forthcoming).
104. Mummolo J, Peterson E. Improving the interpretation of fixed effects regression results. *Polit Sci Res Methods* 2018; **6**: 829–35.

105. Healey J. China is a cyber victim, too. Foreign Policy. 2013. https://foreignpolicy.com/2013/04/16/china-is-a-cyber-victim-too/. (11 March 2022, date last accessed).

106. The White House. Background Press Call by Senior Administration Officials on Malicious Cyber Activity Attributable to the People's Republic of China. https://www.whitehouse.gov/briefing-room/press-briefings/2021/07/19/background-press-call-by-senior-administration-officials-on-malicious-cyber-activity-attributable-to-the-peoples-republic-of-china/. (19 July 2021, date last accessed).

107. Exclusive: China Captures Powerful US NSA Cyberspy Tool. https://www.globaltimes.cn/page/202203/1254856.shtml. (14 March 2022, date last accessed).

108. Ministry of Foreign Affairs of the People's Republic of China. Spokesperson's Remarks on Chinese Cybersecurity Company 360's Technical Report on the Most Powerful Cyber Weapon Used by the US NSA (2022-3-24). https://www.fmprc.gov.cn/mfa_eng/wjb_663304/zzjg_663340/jks_665232/jkxw_665234/202203/t20220329_10656956.html. (24 March 2022, date last accessed).

109. Hollis DB. Why states need an international law for information operations. *Lewis & Clark L. Rev.* 2007; **11**: 1023.

110. Healey J, Maurer T. What it'll take to forge peace in cyberspace. Boston: Christian Science Monitor, 2017.

111. Finnemore M, Hollis DB. Constructing norms for global cybersecurity. *Am J Int Law* 2016; **110**: 425–79.

112. Maurer T. Cyber norm emergence at the United Nations—an analysis of the UN's activities regarding cyber-security. Cambridge: Science, Technology, and Public Policy Program, Belfer Center, 2011.

113. Hare FB. Privateering in cyberspace: should patriotic hacking be promoted as national policy?. *Asian Secur* 2019; **15**: 93–102. doi: 10.1080/14799855.2017.1414803.

114. Egloff FJ. Cybersecurity and non-state actors: a historical analogy with mercantile companies, privateers, and pirates. Oxford: University of Oxford, 2018. https://ora.ox.ac.uk/objects/uuid:77eb9bad-ca00-48b3-abcf-d284c6d27571. (11 March 2022, date last accessed).

115. Sandler T, Siqueira K. Global terrorism: deterrence versus preemption. *Can J Econ Revue canadienne d'économique* 2006; **39**: 1370–87. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5982.2006.00393.x. (11 March 2022, date last accessed).

116. Crosston M. Virtual patriots and a new American cyber strategy: changing the zero-sum game. *Strateg Stud Quart* 2012; **6**: 100–18.

117. MacKinnon R. China's networked authoritarianism. *J Democra* 2011; **22**: 32–46.

118. Hjortdal M. China's use of cyber warfare: espionage meets strategic deterrence. *J Strateg Secur* 2011; **4**: 1–24.

119. Freeman M. The sources of terrorist financing: theory and typology. *Stud Conflict Terror* 2011; **34**: 461–75. doi: 10.1080/1057610X.2011.571193.

120. DeVore MR. Exploring the Iran-Hezbollah relationship: a case study of how state sponsorship affects terrorist group decision-making. *Perspect Terror* 2012; **6**: 85–107.

121. Herr T, Laudrain APB, Smeets M. Mapping the known unknowns of cybersecurity education: a review of syllabi on cyber conflict and security. *J Polit Sci Educ* 2020; 1–17. doi: 10.1080/15512169.2020.1729166.

122. Gorwa R, Smeets M. Cyber conflict in political science: a review of methods and literature. 2019 ISA Working Paper, Toronto: International Studies Association, 2019.

123. Dunn Cavelty Myriam, Elgin Brunner. Introduction: Information, Power, and Security: An Outline of Debates and Implications. *Power and Security in the Information Age* 2007; 1–18.

124. Dunn Cavelty Myriam. Cyber-Security and Private Actors. In Routledge Handbook of Private Security Studies. 2016. https://www.routledge.com/Routledge-Handbook-of-Private-Security-Studies/Abrahamsen-Leander/p/book/9780815347569.

125. Dunn Cavelty Myriam, Reimer A Van Der Vlugt, A Tale of Two Cities: Or How the Wrong Metaphors Lead to Less Security. *Geo. J. Int'l Aff.* 2015; **16**: 2.