Beyond Machines: Humans in Cyber Operations, Espionage, and Conflict[1]

David Danks and Joseph H. Danks

1. The Importance of the Human

It is the height of banality to observe that people, not bullets, fight kinetic wars. The machinery of kinetic warfare is obviously relevant to the conduct of each particular act of warfare, but the reasons for, and meanings of, those acts depend critically on the fact that they are done by humans. Any attempt to understand warfare—its causes, strategies, legitimacy, dynamics, and resolutions—must incorporate humans as an intrinsic part, both descriptively and normatively. Humans from general staff to "boots on the ground" play key roles in all aspects of kinetic warfare, and the literature about it reflects this focus (e.g., the emphasis on understanding the adversary's goals and constraints when developing battle plans).[2] In contrast, many discussions of cyberwarfare and cyber-conflict focus principally on the technical aspects of machines, systems, and data,[3] and human agents are included only as collateral effects (e.g., in discussions about the impact of disabling an

[1] Thanks to the editors, as well as Susannah Paletz and Alan Mishler (both at CASL), for their valuable comments on an earlier version of this chapter. DD was partially supported by a James S. McDonnell Foundation Scholar Award. JHD's work was supported, in whole or in part, with funding from the United States Government. Any opinions, findings and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the view of the University of Maryland, College Park, and/or any agency or entity of the United States Government.

[2] Examples range from the exhortations to know both oneself and one's enemies in Sun Tzu's *The Art of War*, to the emphasis on emotions and other motivations in Clausewitz's *On War*, to quite contemporary work, such as the detailed cognitive analyses of military decision-makers in many chapters of Caroline E. Zsambok and Gary Klein, *Naturalistic Decision Making*, (New York: Psychology Press, 1997).

[3] Examples can be found in many of the papers cited below, as well as Paul Cornish et al., *On Cyber Warfare*, (London: Chatham House, 2010); Paul Rosenzweig, *Cyber Warfare: How Conflicts in Cyberspace Are Challenging America and Changing the World*, (Santa Barbara, CA: Praeger, 2013); Jason Andress and Steve Winterfeld, *Cyber Warfare: Techniques, Tactics and Tools for Security Practitioners*, (Waltham, MA: Syngress, 2011). It should be noted that much of the present volume is a welcome exception to this trend.

adversary's electrical grid), or as loci of moral responsibility (e.g., providing the ground for the moral justification of a cyber-attack).

In some respects, this technical focus is unsurprising: many cyber-capabilities are completely novel in the history of warfare, and it is only natural to focus on the new and original. But this focus has come at a significant cost, as these debates have largely ignored the fact that cyber-actions are typically designed, initiated, and responded to by fallible, cognitively bounded human agents. In particular, the humans that actually engage in cyberwarfare and cyber-conflict are not what many analyses assume—perfectly rational, fully self-aware agents who can automatically bear true moral responsibility, whether praise or blame, for their actions. Instead, cyberwarfare and cyber-operations are conducted by human agents who suffer from cognitive constraints and biases; are often unaware of their own beliefs and desires (present or future); and frequently exhibit failings that undermine their ability to be full moral agents. As we show below, our understanding of the ethics, conduct, and performance of cyberwarfare and cyber-operations changes when we bring the humans back "in the loop."

There are at least four different, morally salient roles for which we should attend to the cognitive limits and features of human agents engaged in cyberwarfare and cyber-operations. Brief examples may help to see the importance of considering cognitively realistic human agents in our analyses. First, humans are the developers of cyber-actions, whether attacks or exploits, and so are arguably responsible for foreseeable outcomes of those actions. As we discuss in § 3, however, the complexity of cyber-systems will frequently exceed our cognitive abilities.[4] In those cases, we will frequently not know the likely outcomes of our actions, even though we have (or rather *should* have, given the large psychological literature on the topic; see citations in § 3) the *meta*-knowledge that we do not know. As a result, we are arguably culpable (at least somewhat) for our ignorance, which thereby reduces or eliminates the mitigating moral power of our ignorance. Our cognitive limitations, coupled with our knowledge of those very limits, potentially imply that many cyber-actions have uncertain moral legitimacy.

Second, human agents are the targets of cyber-actions. The essential humanity of the targets is relevant, for example, in thinking about whether a particular cyber-action is a cyber-*attack*. The definition of a "cyber-attack" proposed in the Tallinn Manual[5] in Rule 30 refers in part to "damage…to objects," which is only a helpful definition if all relevant parties have a shared understanding of what counts as "damage." Unfortunately, it is not clear that there is such an understanding; a categorization of some event as "damage" depends partly on the expectations and culturally-shaped perceptions of the putative "target." As just one instance of this dependence (discussed further in § 4), information extraction could be either an illegitimate, damaging intrusion or appropriate dissemination and sharing, depending on the (partly culturally-determined) "ownership" of that information. The very same cyber-action could be an attack or just publicity, depending on the target's culturally-shaped understanding of the relevant information.

Third, human agents defend against cyber-attacks, -intrusions, and -espionage, and the moral legitimacy of a defensive action will depend partly on cognitive factors, such as

[4] See also David Danks and Joseph H. Danks, "The Moral Permissibility of Automated Responses During Cyberwarfare," *Journal of Military Ethics* 12, no. 1 (2013): 18–33.

[5] Michael N. Schmitt, ed., *Tallinn Manual on the International Law Applicable to Cyber Warfare*, (Cambridge: Cambridge University Press, 2013).

the agents' ability to predict the complex sequence of interactions that could result, or the agents' interpretation of the damage due to the attack. For example, a standard ethical principle about the conduct of warfare is that responses should be proportionate to the triggering event, so application of that principle requires an assessment of the intensity (or degree, or amount) of both the attack and the defense. If human agents were all perfectly rational and fully self-aware, then there could arguably be public, shared assessments of outcome damage. People are not such agents, however; perceptions of the costs and valuations of different actions are sensitive to various cognitive and emotional biases. Our evaluations of cyber-defenders' actions must recognize the complex cognitive origins of the beliefs and perceptions that drive their decisions and actions, particularly as those origins can explain potentially unexpected or seemingly irrational choices by defenders.

Fourth, human agents act as third parties who adjudicate disputes, shape (and sometimes form) public opinion, and generally provide many of the constraints and backgrounds against which cyber-conflicts occur. The people who form these third parties are rarely fully rational, fully self-aware, or fully transparent about their values and goals. Any incorporation of third parties into an analysis of cyberwarfare or cyber-operations must therefore understand them not as ideal agents, but rather as human actors with all of their cognitive, conceptual, and cultural biases, tendencies, and shortcomings.

We have already signaled some of the issues that we will address, such as the morally questionable nature of automatic cyber-actions in light of our cognitive limitations and biases (in § 3), and the importance of incorporating cultural and cognitive features in any analysis of the fuzzy cyberwarfare versus cyber-espionage "boundary" (in § 4). We begin in § 2, however, with a simpler case: the importance of including the human in any solution to the attribution problem. Discussions about how to attribute cyber-actions have focused on technical challenges, but actual cyber-attribution inevitably relies heavily on the imputed motives, constraints, goals, and capacities of various potential adversaries. We thus find our first example of the theme of this chapter, as attribution that fails to be sensitive to human cognitive biases and limits can easily be attribution that goes wrong.

We will refer throughout to three case studies—the 2007 distributed denial-of-service (DDoS) attacks in Estonia,[6] the Stuxnet operation against the Iranian nuclear program,[7] and the apparent Chinese cyber-espionage documented in the Mandiant report[8]—but it is important not to become fixated on the details of any particular case. Our more general moral is simply that our understanding of cyberwarfare and cyber-operations must include human agents as people with cognitive failings and foibles, and not only as interchangeable, identity-less "units" or fully rational, fully self-aware idealizations. Moreover, this is *not* simply a manifestation of the general idea that cognitive details can matter. Rather, analyses in the cyber-domain face novel, distinctive issues and challenges if they presuppose a purely rationalist, idealized understanding of the humans making the cyber-decisions. The complexities of real emotion and cognition matter when thinking about not only the physical bullets of kinetic warfare, but also the binary bullets of cyberwarfare, though differently in the two domains.

[6] Stephen Herzog, "Revisiting the Estonian Cyber Attacks: Digital Threats and Multinational Responses," *Journal of Strategic Security* 4, no. 2 (2011): 49–60.

[7] David Kushner, "The Real Story of Stuxnet," *IEEE Spectrum*, February 26, 2013, http://spectrum.ieee.org/telecom/security/the-real-story-of-stuxnet; Ralph Langner, "Stuxnet's Secret Twin," *Foreign Policy*, November 19, 2013.

[8] *APT1: Exposing One of China's Cyber Espionage Units*, (Mandiant Intelligence Center, February 18, 2013), http://intelreport.mandiant.com.

2. Motives Matter

Perhaps the most-discussed challenge in cyber-operations, particularly in cyber-attacks and cyberwarfare, is the attribution problem:[9] how do we attribute responsibility for some action or event? This question obviously arises in the kinetic domain as well, but is particularly challenging in cyber-contexts. Most discussions of the attribution problem have focused on technical questions:[10] What are the technical conditions under which cyber-attacks can be traced back to their source? Can we determine whether an action originated from a machine that had been compromised in various ways? How should we assign responsibility for distributed attacks? And many other questions. The overall theme of our chapter, however, is to instead ask about the human element. In particular, we contend that attribution is not a purely technical matter, but should also be based on information about motives, opportunities, and behavioral patterns. This additional "human information" not only provides a basis for deciding between equiprobable attributions, but can also lead to attributions that are improbable given solely technical information. In this regard, there are many similarities with so-called "signature strikes" by drones, which should (to be legal or legitimate) depend on not just observed (patterns of) behavior, but also knowledge about potential targets' motives and intentions.[11] Similarly, a criminal law conviction (i.e., attribution of a crime to an individual) requires more than simply

[9] For more detailed discussions see, e.g., Randall R. Dipert, "Preventive War and the Epistemological Dimension of the Morality of War," *Journal of Military Ethics* 5, no. 1 (2006): 32–54; Randall R. Dipert, "The Ethics of Cyberwarfare," *Journal of Military Ethics* 9, no. 4 (2010): 384–410; Neil C. Rowe, "The Ethics of Cyberweapons in Warfare," *International Journal of Cyberethics* 1, no. 1 (2009): 20–31.

[10] As just a limited sample, consider David D. Clark and Susan Landau, "Untangling Attribution," *Harvard National Security Journal* 2, no. 2 (2011); Jeffrey Hunker, Bob Hutchinson, and Jonathan Margulies, *Role and Challenges for Sufficient Cyber-Attack Attribution*, (Institute for Information Infrastructure Protection, January 2008), http://www.thei3p.org/docs/publications/350.pdf; Matthew C. Waxman, "Cyber-Attacks and the Use of Force: Back to the Future of Article 2(4)," *Yale Journal of International Law* 36, no. 2 (2011): 421–59.

[11] Kevin Jon Heller, "'One Hell of a Killing Machine': Signature Strikes and International Law," *Journal of International Criminal Justice* 11, no. 1 (2013): 89–119.

demonstration of means and opportunity; one must also incorporate "human information" (e.g., motives, intentions) about the suspect.

We can easily see why human motives and opportunities matter by thinking about attribution as an inference problem. In almost all cyber-situations, our information set fails to determine the responsible party with certainty. Our attribution inferences are noisy and defeasible because of errors and indeterminism in both our observations and people's decisions and actions. Now suppose that we have only technical knowledge about a particular cyber-action *C*. In this case, we can infer only that *C* must be due to some actor or actors who have the technical capabilities to perform *C*, including (perhaps) the ability to access and control particular machines. In some cases, this inference might be sufficient, as it might uniquely identify the actor. More commonly, though, there will be a non-trivial set of actors who could be responsible for *C*. This underdetermination can be resolved by (i) prior information about which actors were *a priori* most likely to want to perform *C*, as well as (ii) further information about the particular motives or intentions of different actors. In both cases, this information is about the human actors—their motives, intentions, biases, tendencies, and predilections. Of course, we could be mistaken in our assessment of others' reasons, intentions, and desires; our prior beliefs and subsequent information-gathering are obviously not infallible. Moreover, we must focus on the agents' actual preferences and intentions, not their (possibly misleading) publicly-stated ones. Nonetheless, the general point stands: correct attribution requires the integration of both technical assessments and judgments about the (potential) human actors for both of the key terms.[12]

This interaction can be seen in two different examples. Consider first the 2007 DDoS attack against Estonia that occurred during protests and riots by Russian nationalists, triggered by the Estonian government moving a Soviet-era statue/memorial. These cyber-attacks targeted multiple Estonian government agencies and the Estonian financial sector, principally using attack techniques in which key servers were overwhelmed by queries and requests from multiple computers, many of which were presumably compromised themselves.[13] Attribution on purely technical grounds was particularly challenging in this case for two different reasons. First, the attack was highly distributed and employed a large number of compromised machines so enormous expense would be required to trace a significant subset of the commands back to their original source(s). Second, the attack required relatively minimal technical skill and capabilities; many different actors—state, private, and even individual—could have designed and executed the attack. As a result, the technical features alone massively underdetermine the source of the attack. Instead, successful attribution requires the incorporation of information about the "humans in the loop," and in particular, knowledge of which actors would be most probable to launch an attack at this time. In this particular case, that knowledge about the relevant humans did not turn out (at least at first) to uniquely identify

[12] More formally-minded readers will undoubtedly notice the conceptual resemblance of this discussion to a Bayesian analysis. In fact, one could straightforwardly construct a full Bayesian model of the attribution problem both to formalize these intuitions, and also to explore the quantitative interactions between the different types of knowledge. We do not provide such a model here due to both space constraints and a desire to focus on conceptual matters. For those interested in the formal details, the key observations are that (i) $P(A_i) \neq P(A_i \mid PH)$ for actors A_i and prior human information PH (i.e., the priors change when we incorporate human information); and (ii) $P(\mathrm{C} \mid A_i, T) \neq P(\mathrm{C} \mid A_i, T, SH)$ for technical information T and subsequent human information SH (i.e., the likelihoods change as well).

[13] Stephen Herzog, "Revisiting the Estonian Cyber Attacks: Digital Threats and Multinational Responses."

the attackers, largely because there were too many plausible candidates. It did, however, significantly narrow the pool of possibilities from the incredibly large list of those with sufficient technical capabilities.

A second example of the importance of human information in attribution arises in the Stuxnet operation. The Stuxnet worm was introduced into Iranian uranium enrichment facilities from outside (perhaps via a USB drive), and then proceeded to map the facilities' internal structures, change the operations of the enrichment centrifuges to slowly damage them, and finally altered the centrifuge readouts so that everything appeared normal.[14] The Iranian nuclear program was set back as a result, though the impact may have been more limited than initially thought.[15] As with the previous example, attribution could not be resolved on purely technical grounds. Obviously, there were many fewer actors who had the technical capability to carry out the attacks; at the very least, significant state backing (whether official or unofficial) was presumably necessary. Nonetheless, multiple actors, all of whom appear to have engaged in many cyber-activities, arguably had the technical knowledge and capabilities to have performed the attack. Technical features of the attack are thus insufficient to answer the attribution challenge. Instead, one must employ information about the relevant humans in order to determine both (i) the likelihood that actors (with the relevant technical capabilities) would engage in a cyber-attack against the Iranian nuclear program; and (ii) for each possible actor, the likelihood that each actor would choose this particular type of attack (given that they do attack). Arguably, both of these factors point towards a (very) small set of probable actors.

[14] David Kushner, "The Real Story of Stuxnet"; Ralph Langner, "Stuxnet's Secret Twin."

[15] David Albright, Paul Brannan, and Christina Walrond, *Stuxnet Malware and Natanz: Update of ISIS December 22, 2010 Report*, (Institute for Science and International Security, February 15, 2011), http://isis-online.org/uploads/isis-reports/documents/stuxnet_update_15Feb2011.pdf.

Moreover, much of the speculation about who performed the Stuxnet attack seems to have employed exactly this type of reasoning and human information; it seems to have been widely, though implicitly, understood that attribution requires thinking about human motives, interests, and reasoning. In particular, much of the public reporting about Stuxnet for the two years after its discovery[16] argued that the U.S. government was likely involved, principally because of imputed or assumed motives.[17]

The importance of "human information" for attribution purposes in both of these examples is so obvious as to proceed almost unnoticed. In fact, although many high-level discussions of the attribution problem focus on purely technical challenges,[18] the practical reporting, speculation, and discussion of attribution for concrete, particular cyber-attacks almost always invoke human factors.[19] Unfortunately, however, the human is often introduced into the discussion in ways that fail to account for the messiness of actual human cognition. In particular, many attribution discussions assume that the relevant actors (on both sides) are rational agents of the type often studied in decision theory, game theory, or (classical) microeconomics. Such agents have full knowledge of their current desires and interests; can accurately predict their future values and goals; and can design and implement a plan of action that maximizes their chances of success. This rationalist view of the agents and actors is often not explicitly articulated, but is implicitly assumed in

[16] Until the publication of David E. Sanger, *Confront and Conceal: Obama's Secret Wars and Surprising Use of American Power*, (New York: Random House, 2012).

[17] William J. Broad and David E. Sanger, "Worm Was Perfect for Sabotaging Centrifuges," *The New York Times*, November 18, 2010; David E. Sanger, "Iran Fights Malware Attacking Computers," *The New York Times*, September 25, 2010.

[18] See, e.g., Scott J. Shackelford and Richard B. Andres, "State Responsibility for Cyber Attacks: Competing Standards for a Growing Problem," *Georgetown Journal of International Law* 42, no. 4 (2011).

[19] E.g., Sanger, "Iran Fights Malware Attacking Computers;" Ellen Nakashima, "Iranian Hackers Are Targeting U.S. Officials Through Social Networks, Report Says," *Washington Post*, May 29, 2014; Herzog, "Revisiting the Estonian Cyber Attacks: Digital Threats and Multinational Responses."

order to attribute cyber-actions. Inferences of the form "Agent A has reason to do action C. C occurred. Therefore, A is the (or a likely) cause of C." depend on exactly this type of (implicit) view of agents: A must be aware of, understand, and appropriately respond to her rational reasons for action for this inference to be justified.

The past forty years of research in cognitive psychology, behavioral economics, and cognitive neuroscience have given us many reasons to doubt all of these assumptions about actual humans, at least at the individual level: people often do not understand their own reasons for action;[20] they have great difficulty in predicting future desires and preferences;[21] and will often act against their own (future) best interests, particularly in situations in which they face significant threats.[22] The (implicit) rationalist assumption in many attribution discussions is simply false, precisely because cyber-actors are real humans with all of their cognitive biases and foibles. Of course, individual ignorance and irrationality is perfectly consistent with group-level or organizational awareness and rationality, and so perhaps the (implicit) rationalist assumption holds of the groups that, in practice, make many of the most important cyber-decisions. The limited empirical research on the rationality of group planning and decision-making suggests, however, that groups are not necessarily any more rational than the individuals that compose them:[23]

[20] See, e.g., the voluminous literature following from: Richard E. Nisbett and Timothy DeCamp Wilson, "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review* 84, no. 3 (1977): 231–59.

[21] George Loewenstein, Ted O'Donoghue, and Matthew Rabin, "Projection Bias in Predicting Future Utility," *Quarterly Journal of Economics* 118, no. 4 (2003): 1209–48.

[22] Many of the key studies in these areas are collected as references in Dan Ariely, *Predictably Irrational: the Hidden Forces That Shape Our Decisions*, (New York: Harper Collins, 2008); Daniel Kahneman, *Thinking, Fast and Slow*, (New York: Farrar, Straus and Giroux, 2011).

[23] This literature spans forty years of research; see, e.g., John P. Campbell, "Individual Versus Group Problem Solving in an Industrial Sample," *Journal of Applied Psychology* 52, no. 3 (1968): 205–10; Anita Williams Woolley et al., "Bringing in the Experts: How Team Composition and Work Strategy Jointly Shape Analytic Effectiveness," *Small Group Research* 39, no. 3 (2008): 352–71.

some groups can (sometimes) act rationally and in their best interests, but much depends on internal group dynamics.

The importance of recognizing the complexity of human cognition, and the ways that attribution can go wrong when based on rationalist assumptions, can be seen in the Estonia case study. One preliminary attribution for the cyber-attacks, including by the Estonian Foreign Minister, was the Russian government. This attribution seems to have been based principally on the belief that the Russian government had been engaged in systematic retaliation against the Estonian government, and a DDoS attack would have been a rational way to increase the pressure. There seems to have been little initial thought about the possibility that Russian nationalists inside Estonia might be responsible, seemingly because such private citizens were assumed to be uninterested in employing such tactics. It now seems likely that the cyber-attacks actually did come from the second group, driven in large part by rage at perceived attacks that they believed were due to their minority status in Estonia.[24] Standard decision-theoretic or rationalist attribution analyses that assume agents are sensible, self-interested actors would arguably miss such an emotion-based motivation that can lead to disproportionate responses. In contrast, an analysis that incorporates a cognitively sophisticated model of actors and decision-makers could be sensitive to the possibility that some agents will act in seemingly irrational (or at least, arational) ways. Successful attribution depends on understanding how the world appears to different agents, all of whom exhibit various biases and idiosyncrasies; a "one size fits all" approach, particularly a strongly rationalist one, will not work for attribution.

[24] Herzog, "Revisiting the Estonian Cyber Attacks: Digital Threats and Multinational Responses."

The moral of this section is not that attribution is impossible, but rather that many of our current discussions about it are incomplete. Proper attribution requires consideration of not only technical features of the cyber-attack, but also human information and factors about the motives, goals, interests, and constraints of the actors who might be responsible. Moreover, incorporation of this human information is a non-trivial task: one cannot simply assume that all actors know their own motives and act (rationally) on that basis. Rather, attribution must be based on the full messiness of the humans, groups, and organizations that could potentially have been responsible for some cyber-attack. Without this additional complexity, one runs significant risk of misattribution, and so the possibility of a host of ethical, legal, and political problems.

3. Bypassing the (Essential) Human

One key difference between kinetic attacks and cyber-attacks is their speed. Kinetic events often unfold over minutes, hours, days, or longer, while extended, significant sequences of cyber-events can occur in less than a second. This difference in timing leads to a difference in human involvement: human decision-makers are almost always part of kinetic decisions or actions that are morally, politically, or psychologically significant, but not necessarily for corresponding cyber-decisions and cyber-actions. Essentially all ethical norms and principles about the conduct of warfare (open or covert) and espionage assume that humans are "in the loop" to make the morally salient decisions: both the *jus ad bellum* and *jus in bello* aspects of just war theory rely on rational actors having the capacity to consider and make decisions in real time. This assumption is simply false in many cyber-contexts: many significant cyber-actions occur without any human endorsing the specific actions in real time.[25] Instead, those actions are the products of decisions made substantially earlier to institute automated responses that are triggered whenever relevant pre-conditions occur (in an appropriate context).[26] That is, the morally relevant decisions often occur substantially prior to the cyber-event itself.[27]

In the 2007 attacks against Estonia, for example, some servers were receiving orders of magnitude more requests per second than they previously had; that type of speed increase cannot plausibly be achieved by humans "in the loop." The decision to launch the attack was presumably made by a human, but the form of that attack could easily change in the short timespan of cyber-actions into something quite different than what the decision-maker anticipated. Or consider instead much of the cyber-espionage described in the Mandiant report.[28] That report alleged that there have been systematic, targeted intrusions into a range of companies, governments, and other organizations by members of China's People's Liberation Army (PLA). The alleged intrusions involved many different initial attacks, malware families, and other techniques in order to establish entry points, extract information, and leave behind backdoors.[29] Although humans were involved in many stages of the process, a certain level of autonomy was required on both sides in order to respond and adapt to threats and defenses with sufficient speed. For example, a standard cyber-defense involves blocking or ignoring all signals from machines that exhibit certain characteristics; the Mandiant report and other cyber-security bulletins provide "attack signatures" for exactly this reason. More precisely, many different programs

[25] See also Robert Sparrow, "Killer Robots," *Journal of Applied Philosophy* 24, no. 1 (February 2007): 62–77.

[26] See also Heather M. Roff, "Deception, Ruse, and Perfidy in Cyberwarfare," in this volume.

[27] There are also kinetic armaments, most notably landmines, where a decision to deploy is made significantly before the arms are triggered. Cyber-weapons are notably different, however; in particular, neither problem we discuss below arises for these kinetic armaments.

[28] *APT1: Exposing One of China's Cyber Espionage Units*.

[29] Ibid.

provide cyber-defense by watching for behavior that matches an attack signature and then automatically adjusting the target machine in prespecified ways (e.g., ignoring further requests from particular IP addresses). On the other side, multiple well-known cyber-attacks require a series of rapid actions by the attacking machine, each of which is dependent upon particular responses from the target. The details are not particularly important here; the key is simply that cyber-attack, cyber-defense, cyber-espionage, and cyberwarfare all require that humans be "*out* of the loop" at key moments in the sequences of events.[30]

Automaticity poses an ethical challenge precisely because we normally think that the machine cannot itself be a locus of moral responsibility. Software is not morally responsible for some outcome (good or bad); the human who designed, implemented, or used the software bears the moral praise or blame. We do not intend this to be a controversial claim, as we take it to be a natural constraint that responsibility for actions must ultimately inhere in a human decision-maker.[31] The natural question is thus: who bears the responsibility for some automatic action *A*? The natural response is: the individual who decided to implement the automatic response.[32] This answer would be quite sensible if people were rational and fully self-knowledgeable. If that background assumption were true, then the spatial/temporal/systems gap between the human decision and automatic machine action would be irrelevant, as we could justifiably infer

[30] Automaticity is arguably also a feature of certain types of kinetic events: so-called "doomsday devices" provide one infamous example; heat-seeking missiles are a more prosaic case. The observations in this section should apply equally well to automatic responses across domains, though the spatiotemporal separation might well be greater in the cyber-domain than in the kinetic one.

[31] Of course, there might be no one responsible for some *outcome*, as it could be due to luck or some unforeseeable external factor.

[32] Sparrow, "Killer Robots" examines a number of possible loci for moral responsibility.

that the decision-maker still endorsed the action at that later spatiotemporal point. The problem, of course, is that humans are neither fully rational nor fully self-aware, and so we can easily have automatic actions—whether cyber-attack, cyber-defense, or cyber-espionage—for which no human is morally responsible, simply because no human endorses that action at that moment.[33] The situation is somewhat analogous to (though different in key ways from) the kinetic case of a tripwire that triggers a response, but where no human actually endorses the particular response at that moment (e.g., because the wire was tripped by a non-combatant).

More specifically, there are at least two challenges from cognitive science to the necessary background assumptions of rationality and full self-awareness. The first is what we have previously called the chain reaction challenge.[34] Cyber-systems involving multiple automatic actions can readily exhibit complex dynamics that exceed human cognitive abilities for reasoning and inference. Even if we have full knowledge of the different cyber-systems involved in some scenario, we will frequently not be able to adequately predict the likely outcomes due to the complex interrelationships among the components, though we might be able to predict the set of possibilities. And of course, we rarely have anything close to full knowledge, particularly in adversarial contexts in which parties are attempting to hide their systems from one another, or for prediction contexts in which our previous learning (about the system) was driven by more specific goals or needs.[35] We are thus unable to act as fully rational agents in making decisions about cyber-systems, precisely because we are unable to predict or understand the likely impacts of our actions. Events

[33] Danks and Danks, "The Moral Permissibility of Automated Responses During Cyberwarfare."

[34] Ibid.

[35] David Danks, *Unifying the Mind: Cognitive Representations as Graphical Models*, (Cambridge, MA: The MIT Press, 2014).

can rapidly escalate or spiral out of control in ways that are predictable for a fully rational agent, but unpredictable for any actual human. As a result, a human decision-maker might endorse, at time *t*, the decision to implement an automatic response *R*, but only because she fails to understand the role that *R* can play in a later sequence of events. As a result, the spatiotemporal gap between decision and implementation can be meaningful, as the decision-maker could acquire new information (about the system dynamics) that lead her to no longer endorse the previous decision. This problem is actually ubiquitous in cyber-contexts, including non-warfare ones: we face a similar difficulty, for example, in assigning responsibility for financial shocks due to surprising (and deleterious) behaviors of automated trading systems in the complex financial cyber-ecosystem.[36]

The second cognitive science challenge is also a prediction problem, but this time about the decision-maker herself. Consider a decision made at time t_1 about a plan that will be put into action at some later time t_2. The decision-maker must predict at t_1 the preferences, goals, and values she will have at t_2. That is, the decision now should involve the predicted later desires. Of course, one's current preferences and goals could lead to a decision now that constrains one's future self to act somewhat against her (future) interests. Leaving aside these complexities, it is clear that one's future desires and goals should matter in making current decisions. The problem is that people are not necessarily good at predicting their own future preferences, beliefs, desires, and attitudes, and so they can make decisions at t_1 based on incorrect beliefs about what they will want at t_2. For example, people tend to act as if they believe that transient features of their current context

[36] Nathaniel Popper, "Flood of Errant Trades Is a Black Eye for Wall Street," *The New York Times*, August 1, 2012.

will persist indefinitely.[37] As a result, the decision at t_1 can lead to later automatic actions or responses that no one endorses at t_2 and, more importantly, would not have been endorsed at t_1 if the decision-maker had known her likely future desires. We have elsewhere called this the future self-projection bias challenge.[38] This challenge often does not arise in the kinetic domain, as the human decision-maker is plausibly close enough (in space and time) to her own future self to make an accurate prediction of her future preferences and desires.

One might object that these so-called "challenges" are not actually problems for the ethics of automatic cyber-responses, but rather modern instances of the old observation that people can act in morally problematic ways because of ignorance. Moreover, the resulting problematic behavior is often excused on the grounds that the agent "could not have known better"; ignorance actually can be an excuse. An important caveat on these excusings, however, is that the agent not be culpable for the ignorance that led to the action,[39] particularly when the action could have significant consequences.[40] That is, ignorance is only an excuse if one also could not have known that one was ignorant. Both of our challenges—our less-than-full rationality impairing predictions of system dynamics, and our less-than-full self-knowledge and self-awareness leading to incorrect predictions about our future desires—are based on substantial and well-known empirical evidence from the cognitive sciences. Given the importance of many decisions about automatic cyber-responses, ignorance is no excuse, and so the challenges maintain their moral force.

[37] For references to this finding and other types of self-projection bias, see Loewenstein, O'Donoghue, and Rabin, "Projection Bias in Predicting Future Utility."

[38] Danks and Danks, "The Moral Permissibility of Automated Responses During Cyberwarfare."

[39] Gideon Rosen, "Culpability and Ignorance," *Proceedings of the Aristotelian Society* 103, no. 1 (2003): 61–84.

[40] Alexander A. Guerrero, "Don't Know, Don't Kill: Moral Ignorance, Culpability, and Caution," *Philosophical Studies* 136, no. 1 (2007): 59–97.

Instead, we must recognize our own limitations, even if that implies more conservative decision-making than we might otherwise expect or prefer. Humans are "out of the loop" for many cyber-actions, but only because they created the loop in the first place. And so the complexities of human cognition matter for the understanding, assessment, and interpretation of even those cyber-actions that appear to bypass human actors.

4. Soldiers, Saboteurs, Snipers, and Spies in Cyber-space

Rules of law such as International Humanitarian Law have been established for, and govern much of, kinetic warfare. The rules for traditional espionage are less well codified, but nations nonetheless recognize certain common practices and guidelines.[41] The question is whether these same rules of law for warfare and uncodified rules for espionage can be easily extended to cyberwarfare and cyber-espionage, particularly since the distinction between cyberwarfare and cyber-espionage is much fuzzier than between warfare and espionage in the traditional kinetic case. The Tallinn Manual,[42] as part of Rule 66 – Cyber Espionage, holds that:

> cyber espionage is defined narrowly as any act undertaken clandestinely or under false pretences that uses cyber capabilities to gather (or attempt to gather) information with the intention of communicating it to the opposing party.... Cyber espionage must be distinguished from computer network exploitation (CNE), which is a doctrinal, as distinct from an international law, concept. CNE often occurs from beyond enemy territory, using remote access operations. (p. 159)

[41] Neil C. Rowe, "Perfidy in Cyberwarfare," in *Routledge Handbook of Ethics and War*, ed. Fritz Allhoff, Nicholas G Evans, and Adam Henschke, (New York: Routledge, 2013), 394–404.; Roff, "Deception, Ruse, and Perfidy in Cyberwarfare."

[42] Schmitt, *Tallinn Manual on the International Law Applicable to Cyber Warfare*.

The distinction between cyberwarfare and cyber-espionage also is codified within the United States as Title 10 (cyberwarfare) and Title 50 (cyber-espionage),[43] but the laws and rules of other countries do not distinguish the two activities so sharply. In the kinetic domains, the two activities are easier to separate based typically on uniformed military versus civilian spies. In cyber-space, in contrast, identifying the actor is not so clear and the same cyber-intrusion could both have a damaging effect on the target computer network, but also be more benign surveillance.

Normally, the side being attacked can recognize the attacking military by their uniforms and insignia.[44] In cyber-attacks, targets must instead look for other indicators of who the attackers are (i.e., the attribution problem) and whether they are military or civilians. In the Estonian DDoS attack, it remains uncertain as to whether the attack was driven by Russian military, dissident Estonian civilians of Russian background, or some other force. More generally, perfidy is normally not acceptable in warfare lest it become espionage, but perfidy seems inherent in both cyberwarfare and cyber-espionage. Unlike in the kinetic domain, recognizing "cyboteurs" is not at all obvious.[45]

The difficulty in distinguishing between cyber-espionage, cyber-attack, and cyber-defense, but also the importance of doing so, is starkly evident in the increasingly heated accusations between China and the U.S. about various cyber-intrusions.[46] The Mandiant

[43] Andru E. Wall, "Demystifying the Title 10-Title 50 Debate: Distinguishing Military Operations, Intelligence Activities & Covert Action," *Harvard National Security Journal* 3, no. 1 (2011).

[44] But not always, as shown by the confusion in eastern Ukraine during 2014 about whether "rebels" were Ukrainian or Russian, as discussed in, e.g., Andrew Roth and Sabrina Tavernise, "Russians Revealed Among Ukraine Fighters," *The New York Times*, May 27, 2014.

[45] John Arquilla and David Ronfeldt, "The Advent of Netwar (Revisited)," in *Networks and Netwars: the Future of Terror, Crime, and Militancy*, ed. John Arquilla and David Ronfeldt, (RAND Corporation, 2001), 1–25, http://www.rand.org/pubs/monograph_reports/MR1382.html.

[46] David E. Sanger, David Barboza, and Nicole Perlroth, "Chinese Army Unit Is Seen as Tied to Hacking Against U.S.," *The New York Times*, February 19, 2013; David E. Sanger and Nicole Perlroth, "Hackers From China

report attributed numerous cyber-attacks to China's People's Liberation Army (PLA) 2nd Bureau, 3rd General Staff Department Unit 61398 in Shanghai, and presented substantial evidence supporting this attribution.[47] Over many years, the active exploits (that Mandiant attributes to the Shanghai group) have become legion, involving the exfiltration of massive amounts of intellectual property from U.S. businesses, including defense contractors and high-profile security firms such as RSA Security LLC.[48] The attribution of the exploitation to Unit 61398 helped lead to the indictments of five members of the P.L.A.—Wang Dong, Sun Kailiang, Wen Xinyu, Huang Zhenyu, and Gu Chunhui, all officers in Unit 61398, all pictured in uniforms of the PLA—by a grand jury in Pittsburgh.[49] These five were accused of hacking into computer networks operated by several U.S. businesses operating critical infrastructure in western Pennsylvania, such as Westinghouse, U.S. Steel, SolarWorld, Alcoa, and United Steel Workers.

One key, largely unanswered question, is whether the indicted officers of Unit 61398 should be viewed as combatants (soldiers) or as spies? Is it relevant that the indictments included photographs of the accused in P.L.A. uniforms? Standard international laws of war would contend that it is conceivably relevant, as the agreements governing soldiers are very different than the national laws applicable to spies operating within a country, but that presupposes that simply wearing a uniform matters (in these ways) in cyber-space. More generally, these diverse, but linked, roles—non-combatant vs.

Resume Attacks on U.S. Targets," *The New York Times*, May 19, 2013; David E. Sanger, "With Spy Charges, U.S. Draws a Line That Few Others Recognize," *The New York Times*, May 19, 2014.

[47] See citations earlier in § 3.

[48] RSA Security LLC is an American computer and network security firm known for the RSA public-key cryptography algorithm.

[49] Office of Public Affairs, U S Department of Justice, "U.S. Charges Five Chinese Military Hackers for Cyber Espionage Against U.S. Corporations and a Labor Organization for Commercial Advantage," May 19, 2014, http://www.justice.gov/opa/pr/2014/May/14-ag-528.html; Michael S. Schmidt and David E. Sanger, "5 in China Army Face U.S. Charges of Cyberattacks," *The New York Times*, May 19, 2014.

military vs. support vs. spy—are partly culturally constructed and constituted, and so our analysis of them must include information about the cognitively bounded humans involved. Even if the officers were moonlighting in their off-duty hours as civilians and outside their employment in the military, their activities are just as elusive and difficult to prove in a court of law, either national or international.

Moreover, the Chinese government vigorously denied this attribution and attacked the motives of the U.S. government in pursuing the indictments.[50] The U.S. government and military's hands are arguably far from clean, as the Chinese and other countries are wont to point out: "China is a victim of severe US cyber theft, wiretapping and surveillance activities. Large amounts of publicly disclosed information show that relevant US institutions have been conducting cyber intrusion, wiretapping and surveillance activities against Chinese government departments, institutions, companies, universities and individuals."[51]

A hypothetical case can help to draw out some of the complexities that can arise in this fuzzy area between cyberwarfare and cyber-espionage. Suppose an agency of the fictional Yahere government infiltrated firms in the country of Nowhere; let ChatMor be one such company, a leading computer communications firm with a worldwide reach. Further suppose that the country of Nowhere is considered to be an ideological, political, economic, and military challenger to Yahere. Successful access to ChatMor's computer networks to exfiltrate its intellectual property would also allow Yahere's government agencies to spy not only on Nowhere's departments and institutions, but also on

[50] Ministry of Foreign Affairs of the People's Republic of China, "China Reacts Strongly to U.S. Announcement of Indictment Against Chinese Personnel," May 19, 2014, http://www.fmprc.gov.cn/mfa_eng/xwfw_665399/s2510_665401/t1157487.shtml.

[51] Ibid.

organizations in other countries that purchased and installed ChatMor's products. Whether such actions were taken by Yahere's military or civilian personnel is largely irrelevant when the actions and end use is for espionage and reconnaissance. But when the knowledge of ChatMor's network operations might be exploited for cyber-attacks (i.e., cyberwarfare), then the legal status arguably changes under international law.

These observations about the fuzzy boundary between cyberwarfare and cyber-espionage all draw on the importance of perceptions, whether about the cyber-action, the effects of the action, or the roles of the people involved. These kinds of perceptions depend in key ways on the cultural frames and biases that we bring to bear. Whether a particular cyber-action is defense or stealing is in the eye of the actor or the target. For example, the Chinese and Americans have quite different views of cyber-espionage, at least when motivated by defense or economic exploitation. There is no international law governing economic espionage.[52] The Chinese position, as illustrated in the previous quotation from the PRC Ministry of Foreign Affairs, conflates activities of governments, especially actions related to national defense, with economic activities of private enterprise. Defending the nation and defending the economy are literally one and the same. In contrast, the U.S. distinguishes quite sharply between government and private enterprise. As U.S. Attorney General Eric Holder asserted in announcing the indictments:

> The range of trade secrets and other sensitive business information stolen in this case is significant and demands an aggressive response. Success in the global market place should be based solely on a company's ability to innovate and compete, not on a sponsor government's ability to spy and steal business secrets.[53]

[52] David P. Fidler, "Economic Cyber Espionage and International Law: Controversies Involving Government Acquisition of Trade Secrets Through Cyber Technologies," *ASIL Insights* 17, no. 10 (March 20, 2013).

[53] Office of Public Affairs, U S Department of Justice, "U.S. Charges Five Chinese Military Hackers for Cyber Espionage Against U.S. Corporations and a Labor Organization for Commercial Advantage."

These sorts of alleged activities by both U.S. and Chinese governments, whether by uniformed military or civilians, blur the lines between cyberwarfare, cyber-espionage, and cyber-defense.[54] Technical specifications of the activities are not at issue. Rather, the human motives, perceptions, social, political, and economic perspectives are the issue.

Human motives are neither simple nor pure. One might initially attribute cyber-activities of military personnel to nationalistic or patriotic motives, or even ideology, but hackers are also motivated by money, power, or success. Economic gain is a frequently attributed extrinsic motivator since it can be observed and measured. But hackers' motives, like those of all of us, are multiply determined. The need to gain power over an individual, organization, or target can be a powerful motivator for the psychological satisfaction it yields. Hackers by their practice are good at solving problems, especially the technical problems of computers and computer networks, and this general problem-solving orientation can lead them to use that skill to achieve power and control over computer networks. Hackers are social beings as well and can be motivated by the need to affiliate with fellow hackers or to be loyal to their organizations or country. They also may desire recognition from their peers. The prevalence of "black hat" organizations attests to the strength of several of these motives.[55]

The Stuxnet virus that attacked the uranium centrifuges in Iran is another instance of the blurring of the line between cyberwarfare and cyber-espionage, as there were both kinetic and cyber effects. The Stuxnet worm has been commonly attributed to the U.S. and

[54] Sanger, "With Spy Charges, U.S. Draws a Line That Few Others Recognize"; Sanger, David E., and Nicole Perlroth. "U.S. Penetrated Chinese Servers It Saw as Spy Risk." *The New York Times.* March 23, 2014.

[55] Dorothy E. Denning, "Cyber Conflict as an Emergent Social Phenomenon," in *Corporate Hacking and Technology-Driven Crime: Social Dynamics and Implications*, ed. T J Holt and B H Schell, (Hershey, NY: Information Science Reference, 2011), 170–86; M. K. Rogers, "The Psyche of Cybercriminals: a Psychosocial Perspective," in *Cybercrimes: a Multidisciplinary Analysis*, ed. Sumit Ghosh and Elliott Turini, (Springer eBooks, 2011), 217–35, http://link.springer.com/book/10.1007/978-3-642-13547-7.

Israel although neither government has publicly acknowledged responsibility for it. The physical damage to the centrifuges has been well publicized, but what is less well known is the cyber-effect of the Stuxnet virus that was intended to hide the physical attack. Specifically, Stuxnet altered the information feedback to the operators so that the centrifuges appeared to be operating normally. As a result, the operators were unaware that the centrifuges were speeding up and out of control.[56]

A possible Iranian response to Stuxnet has recently been revealed by iSightPartners.[57] A social networking campaign, dubbed NEWSCASTER by iSightPartners, targeted U.S. military officers and diplomats as well as both U.S. and Israeli defense contractor personnel. The hackers invented numerous on-line personae in various social networking sites such as Facebook and LinkedIn, and then sent friendly messages with known contacts of their targets, including links to a fake site, NewsOnAir.org, a technique known as spearphishing.[58] (NewsOnAir.com is a legitimate English-language news site in India, so the ruse was plausible.) Once the target's friends and contacts had been compromised, the targets themselves could be spearphished. Once the primary target had been compromised, data and information could be exfiltrated for a variety of adverse uses. NEWSCASTER is a well-defined social engineering campaign that relies on human attackers and spies being able to take advantage of the cognitive and emotional biases and weaknesses of the human targets.

[56] Kushner, "The Real Story of Stuxnet;" Thomas Rid, *Cyber War Will Not Take Place*, (Oxford: Oxford University Press, 2013).

[57] iSightPartners, *NEWSCASTER: an Iranian Threat Within Social Networks*, May 28, 2014.

[58] Phishing is a social engineering technique to obtain private information. The attacker sends an e-mail message that appears to be from a trusted source requesting private information, such as passwords and bank account numbers. Phishing attacks are sent to a large number of targets without personal links to the individual target. Spearphishing attacks also appear to come from a trusted source, often from within the recipient's own organization or a website the target uses frequently, and are targeted to particular individuals using personal targeting information.

Successful cyber-attacks and cyber-espionage, and even cyber-defense, require a full understanding of the human target, as well as implicit self-knowledge of the attacking operator. An analysis of specific instances of cyber-attacks and cyber-espionage reveals this human-centric character, but cyber-security experts rarely, if ever, talk about the human element, except for campaigns to get computer users to improve their security activities. Effective cyber-attacks and cyber–espionage frequently depend on social engineering, including manipulating people psychologically to get them to take actions they might not otherwise take, especially with regard to confidential information. These manipulations are based on understanding the biases and weaknesses of human decision making; these cognitive foibles might be called bugs in the human wetware, but are inevitable and perhaps even necessary. Understanding the cognitive, behavioral, and social processes of the humans in the loop is crucial when exploring the distinctions between cyberwarfare and cyber-espionage across a range of cultural contexts.

5. Keeping Humans in the Picture

Humans—not computers, networks, or technology writ large—are the agents who can be held responsible for ethical decisions. Key issues in the ethics of cyberwarfare depend critically on recognizing and understanding these agents as cognitively real humans, rather than idealized fictions. Humans must be kept "in the loop" to decide: who is the aggressor; whether an action discriminates between effects on combatants and civilians; if an action is a proportional response; whether combatants are clearly identified as legitimate targets; or if there can even be perfidy or treacherous deceit in cyber-space.[59] Technical information about a cyber-attack is insufficient to answer these questions, as was demonstrated in the DDoS attacks on Estonia. In criminal activity, the key elements in identifying a perpetrator are whether he has the motivation, opportunity, and capability to commit the crime. The same is true for attributing a cyber-attack or cyber-espionage to an agent: we must determine which individual, organization, or nation has the technical capability to actually effect the cyber-attack or –espionage, and just as importantly, the motivation to carry out a cyber-attack and the opportunity to do so. Technical information is helpful for these determinations, but not sufficient; we also must think about the actual humans behind the cyber-actions.

The speed and velocity at which cyber-actions take place is a fundamental challenge for human cyber-operators. Automatic responses leave humans outside of the real-time loop. Humans can be held responsible for developing and programming the automatic responses, but it is unclear under what conditions the responsibility for the actual event transfers to those agents. If humans were rational in their planning and decisions, then assigning responsibility to originating agents might be reasonable, but neither people nor groups exhibit the necessary rationality or self-knowledge. In particular, people are frequently unable to accurately predict their future preferences, goals, and values. Moreover, preferences for a particular course of action often are modified by changes in the physical situation as well as the person's internal mental state.

[59] Patrick Lin, Fritz Allhoff, and Neil C. Rowe, "War 2.0: Cyberweapons and Ethics," *Communications of the ACM* 55, no. 3 (March 2012): 24–26.

In addition to not recognizing their own (future) motivations, people frequently misjudge others' motivations, a finding often called the fundamental attribution error.[60] Our own motives are commonly attributed to transitory situational factors at the moment when the decision is made and action taken, a stance that is sometimes called "situationism." The motivations of others, however, are typically attributed to relatively enduring (personality) traits, that is, a stance of "dispositionism." These persistent personality traits are more unidimensional, and so attributing actions to them can result in simplistic, and often erroneous, interpretations of another's actions, especially when that person is a potential adversary. Moreover, the fundamental attribution error is culturally bound. It is not found as frequently in groups from so-called collectivistic cultures (e.g., East Asian) as opposed to groups in individualistic cultures (e.g., Western European and American).[61] This cultural differentiation results from East Asians holding a more situational perspective than Westerners: "East Asians believe dispositions to be more malleable and have a more holistic conception of the person as being situated in a broad social context."[62] If we instead recognize other humans as complex cognitive agents, then we can consider their situational context as well as their enduring personality traits, and thereby recognize and conceive of multiple motives.

All of these factors point to the need for a human-centric analysis of the cyber-actors and the cyber-events. Only with a full understanding of the human in the loop can we begin

[60] Edward E. Jones and Victor A. Harris, "The Attribution of Attitudes," *Journal of Experimental Social Psychology* 3, no. 1 (1967): 1–24; Richard E. Nisbett and Lee Ross, *The Person and the Situation: Perspectives of Social Psychology*, (New York: McGraw-Hill Publishing, 1991).

[61] J. G. Miller, "Culture and the Development of Everyday Social Explanation," *Journal of Personality and Social Psychology* 46, no. 5 (1984): 961–78.

[62] p. 49 of Incheol Choi, Richard E. Nisbett, and Ara Norenzayan, "Causal Attribution Across Cultures: Variation and Universality," *Psychological Bulletin* 125 (1999): 47–63; see also Ara Norenzayan, Incheol Choi, and Richard E. Nisbett, "Cultural Similarities and Differences in Social Inference: Evidence From Behavioral Predictions and Lay Theories of Behavior," *Personality and Social Psychology Bulletin* 28 (2002): 109–20.

to assign and understand ethical responsibility for these actions, and more generally explicate the complex moral and ethical dimensions of cyberwarfare, cyber-espionage, and other actions in the cyber-domain.

Bibliography

Albright, David, Paul Brannan, and Christina Walrond. *Stuxnet Malware and Natanz: Update of ISIS December 22, 2010 Report*, Institute for Science and International Security, February 15, 2011. http://isis-online.org/uploads/isis-reports/documents/stuxnet_update_15Feb2011.pdf.

Andress, Jason, and Steve Winterfeld. *Cyber Warfare: Techniques, Tactics and Tools for Security Practitioners*, Waltham, MA: Syngress, 2011.

APT1: Exposing One of China's Cyber Espionage Units, Mandiant Intelligence Center, February 18, 2013. http://intelreport.mandiant.com.

Ariely, Dan. *Predictably Irrational: the Hidden Forces That Shape Our Decisions*, New York: Harper Collins, 2008.

Arquilla, John, and David Ronfeldt. “The Advent of Netwar (Revisited).” In *Networks and Netwars: the Future of Terror, Crime, and Militancy*, edited by John Arquilla and David Ronfeldt, 1–25, RAND Corporation, 2001. http://www.rand.org/pubs/monograph_reports/MR1382.html.

Broad, William J., and David E. Sanger. “Worm Was Perfect for Sabotaging Centrifuges.” *The New York Times*, November 18, 2010.

Campbell, John P. “Individual Versus Group Problem Solving in an Industrial Sample.” *Journal of Applied Psychology* 52, no. 3 (1968): 205–10.

Choi, Incheol, Richard E. Nisbett, and Ara Norenzayan. “Causal Attribution Across Cultures: Variation and Universality.” *Psychological Bulletin* 125 (1999): 47–63.

Clark, David D., and Susan Landau. “Untangling Attribution.” *Harvard National Security Journal* 2, no. 2 (2011).

Cornish, Paul, David Livingstone, Dave Clemente, and Claire Yorke. *On Cyber Warfare*, London: Chatham House, 2010.

Danks, David. *Unifying the Mind: Cognitive Representations as Graphical Models*, Cambridge, MA: The MIT Press, 2014.

Danks, David, and Joseph H. Danks. “The Moral Permissibility of Automated Responses During Cyberwarfare.” *Journal of Military Ethics* 12, no. 1 (2013): 18–33.

Denning, Dorothy E. “Cyber Conflict as an Emergent Social Phenomenon.” In *Corporate Hacking and Technology-Driven Crime: Social Dynamics and Implications*, edited by T J Holt and B H Schell, 170–86, Hershey, NY: Information Science Reference, 2011.

Dipert, Randall R. “Preventive War and the Epistemological Dimension of the Morality of War.” *Journal of Military Ethics* 5, no. 1 (2006): 32–54.

Dipert, Randall R. “The Ethics of Cyberwarfare.” *Journal of Military Ethics* 9, no. 4 (2010): 384–410.

Fidler, David P. “Economic Cyber Espionage and International Law: Controversies Involving Government Acquisition of Trade Secrets Through Cyber Technologies.” *ASIL Insights* 17, no. 10 (March 20, 2013).

Guerrero, Alexander A. “Don’t Know, Don’t Kill: Moral Ignorance, Culpability, and Caution.” *Philosophical Studies* 136, no. 1 (2007): 59–97.

Heller, Kevin Jon. “‘One Hell of a Killing Machine’: Signature Strikes and International Law.” *Journal of International Criminal Justice* 11, no. 1 (2013): 89–119.

Herzog, Stephen. “Revisiting the Estonian Cyber Attacks: Digital Threats and Multinational Responses.” *Journal of Strategic Security* 4, no. 2 (2011): 49–60.

Hunker, Jeffrey, Bob Hutchinson, and Jonathan Margulies. *Role and Challenges for Sufficient Cyber-Attack Attribution*, Institute for Information Infrastructure Protection, January 2008. http://www.thei3p.org/docs/publications/350.pdf.

iSightPartners. *NEWSCASTER: an Iranian Threat Within Social Networks*, May 28, 2014.

Jones, Edward E., and Victor A. Harris. “The Attribution of Attitudes.” *Journal of Experimental Social Psychology* 3, no. 1 (1967): 1–24.

Kahneman, Daniel. *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux, 2011.

Kushner, David. “The Real Story of Stuxnet.” *IEEE Spectrum*, February 26, 2013. http://spectrum.ieee.org/telecom/security/the-real-story-of-stuxnet.

Langner, Ralph. “Stuxnet's Secret Twin.” *Foreign Policy*, November 19, 2013.

Lin, Patrick, Fritz Allhoff, and Neil C. Rowe. “War 2.0: Cyberweapons and Ethics.” *Communications of the ACM* 55, no. 3 (March 2012): 24–26.

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin. “Projection Bias in Predicting Future Utility.” *Quarterly Journal of Economics* 118, no. 4 (2003): 1209–48.

Miller, J. G. “Culture and the Development of Everyday Social Explanation.” *Journal of Personality and Social Psychology* 46, no. 5 (1984): 961–78.

Ministry of Foreign Affairs of the People's Republic of China. “China Reacts Strongly to U.S. Announcement of Indictment Against Chinese Personnel,” May 19, 2014.

http://www.fmprc.gov.cn/mfa_eng/xwfw_665399/s2510_665401/t1157487.shtml.

Nakashima, Ellen. "Iranian Hackers Are Targeting U.S. Officials Through Social Networks, Report Says." *Washington Post*, May 29, 2014.

Nisbett, Richard E., and Lee Ross. *The Person and the Situation: Perspectives of Social Psychology*, New York: McGraw-Hill Publishing, 1991.

Nisbett, Richard E., and Timothy DeCamp Wilson. "Telling More Than We Can Know: Verbal Reports on Mental Processes." *Psychological Review* 84, no. 3 (May 1977): 231–59.

Norenzayan, Ara, Incheol Choi, and Richard E. Nisbett. "Cultural Similarities and Differences in Social Inference: Evidence From Behavioral Predictions and Lay Theories of Behavior." *Personality and Social Psychology Bulletin* 28 (2002): 109–20.

Office of Public Affairs, U S Department of Justice. "U.S. Charges Five Chinese Military Hackers for Cyber Espionage Against U.S. Corporations and a Labor Organization for Commercial Advantage," May 19, 2014. http://www.justice.gov/opa/pr/2014/May/14-ag-528.html.

Popper, Nathaniel. "Flood of Errant Trades Is a Black Eye for Wall Street." *The New York Times*, August 1, 2012.

Rid, Thomas. *Cyber War Will Not Take Place*, Oxford: Oxford University Press, 2013.

Roff, Heather M. "Deception, Ruse, and Perfidy in Cyberwarfare," this volume.

Rogers, M. K. "The Psyche of Cybercriminals: a Psychosocial Perspective." In *Cybercrimes: a Multidisciplinary Analysis*, edited by Sumit Ghosh and Elliott Turini, 217–35, Springer eBooks, 2011. http://link.springer.com/book/10.1007/978-3-642-13547-7.

Rosen, Gideon. "Culpability and Ignorance." *Proceedings of the Aristotelian Society* 103, no. 1 (2003): 61–84.

Rosenzweig, Paul. *Cyber Warfare: How Conflicts in Cyberspace Are Challenging America and Changing the World*, Santa Barbara, CA: Praeger, 2013.

Roth, Andrew, and Sabrina Tavernise. "Russians Revealed Among Ukraine Fighters." *The New York Times*, May 27, 2014.

Rowe, Neil C. "Perfidy in Cyberwarfare." In *Routledge Handbook of Ethics and War*, edited by Fritz Allhoff, Nicholas G. Evans, and Adam Henschke, 394–404, New York: Routledge, 2013.

Rowe, Neil C. "The Ethics of Cyberweapons in Warfare." *International Journal of Cyberethics* 1, no. 1 (2009): 20–31.

Sanger, David E. *Confront and Conceal: Obama's Secret Wars and Surprising Use of American Power*, New York: Random House, 2012.

Sanger, David E. "Iran Fights Malware Attacking Computers." *The New York Times*, September 25, 2010.

Sanger, David E. "With Spy Charges, U.S. Draws a Line That Few Others Recognize." *The New York Times*, May 19, 2014.

Sanger, David E., and Nicole Perlroth. "Hackers From China Resume Attacks on U.S. Targets." *The New York Times*, May 19, 2013.

Sanger, David E., and Nicole Perlroth. "U.S. Penetrated Chinese Servers It Saw as Spy Risk." *The New York Times.* March 23, 2014.

Sanger, David E., David Barboza, and Nicole Perlroth. "Chinese Army Unit Is Seen as Tied to Hacking Against U.S." *The New York Times*, February 19, 2013.

Schmidt, Michael S., and David E. Sanger. "5 in China Army Face U.S. Charges of Cyberattacks." *The New York Times*, May 19, 2014.

Tallinn Manual on the International Law Applicable to Cyber Warfare. Edited by Michael N Schmitt, Cambridge: Cambridge University Press, 2013.

Shackelford, Scott J., and Richard B. Andres. "State Responsibility for Cyber Attacks: Competing Standards for a Growing Problem." *Georgetown Journal of International Law* 42, no. 4 (2011).

Sparrow, Robert. "Killer Robots." *Journal of Applied Philosophy* 24, no. 1 (February 2007): 62–77.

Wall, Andru E. "Demystifying the Title 10-Title 50 Debate: Distinguishing Military Operations, Intelligence Activities & Covert Action." *Harvard National Security Journal* 3, no. 1 (2011).

Waxman, Matthew C. "Cyber-Attacks and the Use of Force: Back to the Future of Article 2(4)." *Yale Journal of International Law* 36, no. 2 (2011): 421–59.

Woolley, Anita Williams, Margaret E. Gerbasi, Christopher F. Chabris, Stephen M. Kosslyn, and J. Richard Hackman. "Bringing in the Experts: How Team Composition and Work Strategy Jointly Shape Analytic Effectiveness." *Small Group Research* 39, no. 3 (2008): 352–71.

Zsambok, Caroline E., and Gary Klein. *Naturalistic Decision Making*, New York: Psychology Press, 1997.