Big Data and Bad Data: On the Sensitivity of Security Policy to Imperfect Information

James T. Graves,† Alessandro Acquisti†† & Nicolas Christin‡

On two occasions I have been asked,—"Pray, Mr. Babbage, if you put into the machine wrong figures, will the right answers come out?" . . . I am not able rightly to apprehend the kind of confusion of ideas that could provoke such a question.

Charles Babbage[1]

In this Essay, we examine some of the factors that make developing a "science of security" a significant research and policy challenge. We focus on how the empirical hurdles of missing data, inaccurate data, and invalid inferences can significantly impact—and sometimes impair—the security decisionmaking processes of individuals, firms, and policymakers. We offer practical examples of the sensitivity of policy modeling to those hurdles and highlight the relevance of these examples in the context of national security.

INTRODUCTION

In recent years, policy and research circles have directed growing attention toward the goal of developing a "science of security." In this Essay, we consider the empirical challenges of developing such a science. We highlight how imperfect information affects the way security trade-offs are measured, and we

† PhD Candidate (Engineering and Public Policy) 2016, Carnegie Mellon University.

†† Professor of Information Technology and Public Policy, Carnegie Mellon University.

‡ Assistant Research Professor of Electrical and Computer Engineering, Carnegie Mellon University.

This work was partially funded by the Department of Homeland Security Science and Technology Directorate, Cyber Security Division, Broad Agency Announcement 11.02; the Government of Australia; and SPAWAR Systems Center Pacific, via contract number N66001-13-C-0131. Portions of this work were also supported by NSF IGERT grant DGE-0903659. In addition, Acquisti gratefully acknowledges support from the Carnegie Corporation of New York via an Andrew Carnegie Fellowship. This Essay represents the position of the authors and not that of the aforementioned agencies.

1 Charles Babbage, *Passages from the Life of a Philosopher* 67 (Longman 1864) (quotation marks omitted).

discuss how, based on those potentially inaccurate measurements, different stakeholders can end up making poor security decisions.

In 2011, the White House published a strategic plan for cybersecurity research and development.[2] As one of four thrusts, the plan emphasized the need to develop a science of security that would formalize rigorous principles and produce fundamental building blocks to design secure and trustworthy information systems.[3] Secure information systems, in turn, would protect not only individual firms and consumers but also national interests against espionage, terrorism, and cyberwars. The plan supported a call that had grown louder over the years within the security research community: the need to move from security as merely "engineering" to security as "science." Security engineering practices are often deemed to be ad hoc and reactive: "find a bug; patch it; find the next bug; and so on."[4] However, a truly secure system should be able to defend against any and all possible attacks, including attacks that are not yet known to—or foreseeable by—the defender.[5] To develop such systems, a *science* of cybersecurity is needed. Such a science would consist of a body of security laws and first principles that transcend specific technologies and systems. A science of cybersecurity should also be able to produce models and abstractions that are amenable to rigorous treatments, experimentation, replication, and—eventually—generalizable solutions.

Since the publication of the White House's cybersecurity plan, efforts in this area by the research community and several funding agencies (such as the National Science Foundation, the National Security Agency, and the Defense Advanced Research Projects Agency) have been gaining momentum.[6] Yet the challenges that

---

2 See generally National Science and Technology Council, *Trustworthy Cyberspace: Strategic Plan for the Federal Cybersecurity Research and Development Program* (Executive Office of the President, Dec 2011), archived at http://perma.cc/W456-7LGH.

3 See id at *1 ("The priorities are organized into four thrusts: Inducing Change, Developing Scientific Foundations, Maximizing Research Impact, and Accelerating Transition to Practice.").

4 Munindar P. Singh, *Toward a Science of Security* (Computing Now, Jan 2013), archived at http://perma.cc/3ED4-WGYZ.

5 See Fred B. Schneider, *Blueprint for a Science of Cybersecurity*, 19 Next Wave 47, 47 (2012), archived at http://perma.cc/CMH7-QNJL.

6 Frederick R. Chang, *Guest Editor's Column*, 19 Next Wave i, i (2012), archived at http://perma.cc/Q4KV-LYXF ("There are some promising indications that a science of cybersecurity initiative is gaining momentum, including several workshops, conferences, and reports that point to the need for an interdisciplinary approach to addressing the

arise in the attempt to make security a science are numerous. A science of security should provide open and evolving principles that guarantee the trustworthiness and robustness of systems.[7] It should guide the design of services that are resilient to the aggression of attackers whose strategies will be continually adapting.[8] It should be informed by, and weave together, insights from highly diverse disciplines, including computer science, economics, and psychology.[9] In addition, a science of security should be empirically grounded: it should rely on rigorous measurements of security phenomena in order to precisely quantify security trade-offs, and, based on those trade-offs, it should provide guidance for the decisions of different stakeholders.[10]

The aim of this Essay is to examine the challenges associated with the latter goals: how security trade-offs are quantified and then used in the decisionmaking processes of various stakeholders. These stakeholders may be agents in the marketplace (such as firms that are exposed to cyberattacks and consumers who use those firms' services), but they may also be governments: cyberthreat actors increasingly include nation-states that undertake "offensive cyber operations against private sector targets to support their economic and foreign policy objectives."[11] In Part I, we focus on how imperfect information can impact the calculation of those trade-offs and hence swing the decisionmaking of those agents. By "imperfect information," we refer, quite loosely, to a combination of different data challenges that we categorize into three types of problems: missing data, inaccurate data, and invalid inferences.

The first type of problem, missing data, can happen for a number of reasons. The data might not exist at all. The data could exist only in a form that is extremely difficult to access or use. Or the data could exist and be usable but be kept secret. For instance, many organizations in both the private and public

problem."); Robert Meushaw, *NSA Initiatives in Cybersecurity Science*, 19 Next Wave 8, 10–11 (2012), archived at http://perma.cc/Q4KV-LYXF (discussing funding for cybersecurity research).

7 See Singh, *Toward a Science of Security* (cited in note 4).

8 See id.

9 Schneider, *Blueprint* at 53 (cited in note 5).

10 See Charles H. Brown, et al, *The Science of Security: A Survey and Analysis* *2 (AFCEA International Cyber Committee, June 2014), archived at http://perma.cc/39CH-XUJF (highlighting the various areas that are reliant on cybertechnologies and the broad motivations for developing a science of cybersecurity).

11 James R. Clapper, *Worldwide Cyber Threats* *2 (House Permanent Select Committee on Intelligence, Sept 10, 2015), archived at http://perma.cc/7XN6-9N5W.

sectors hold large amounts of data that could help government agencies and researchers better understand the behavior of attackers.[12] But these organizations are often reluctant to share data due to legal concerns, policy concerns, and perceptions of the sensitive or proprietary nature of that data.[13]

The second problem is inaccurate data. The adage "garbage in, garbage out" is nearly as old as computing itself, and examples abound of poor decisions being made because of poor data. Whether due to incomplete data collection, uncertain estimates, "fat-fingered" data entry, miscategorization, or other reasons, the consequences of inaccurate data are clear: if the input is inaccurate, the output will be, too.

The third problem, invalid inferences, occurs when people or organizations make incorrect extrapolations and analyses from available and possibly accurate data. Invalid inferences may be the rational result of decisionmaking under particular incentives or they could result from cognitive biases. For example, perverse incentives can play a role in how agents underestimate or overestimate the trade-offs associated with security incidents. A particularly pervasive example of bias is the assumption that correlation implies causation. The problem of invalid inferences is a problem of analysis, not of data quality—but inaccurate analysis can lead to inaccurate data, which might then be used as if the data were reliable.

These informational problems are certainly not exclusive to the security domain. Yet, as we discuss, they seem particularly pervasive in the cybersecurity realm and can significantly affect crucial decisions by various stakeholders.

In Part I, we offer a series of examples of imperfect information affecting the estimation of security trade-offs and the decisionmaking of various agents, relying heavily on our own ongoing empirical research in this area. Our examples focus mainly on interactions in the private sector. But the empirical hurdles we highlight in this Essay also affect the analysis of the trade-offs associated with threats to national interests, such as the

---

[12] See *Cyberspace Policy Review: Assuring a Trusted and Resilient Information and Communications Infrastructure* *17, archived at http://perma.cc/2J26-2SM7 ("The public and private sectors' interests are intertwined with a shared responsibility for ensuring a secure, reliable infrastructure.").

[13] See id at *18–19 ("Industry has also expressed reservations about disclosing to the Federal government sensitive or proprietary business information, such as vulnerabilities and data or network breaches.").

critical infrastructure. Therefore, in Part II, we comment on the relevance of our analysis to the context of national security.

I. IMPERFECT INFORMATION: BAD DATA, BAD DECISIONS

Imperfect information is often used in public discourse and can potentially impact decision outcomes by individual users, firms, and governments. In this Part, we discuss examples of each type of data problem.

A. Missing Data

The problem of missing data (or, using the language of economics, "incomplete information") is pervasive in information security. For instance, organizations that have fallen victim to cyberattacks may not be keen to share this information with others for fear of renewed attacks, reputational costs, or stock market losses.[14] Thus, a substantial number of security incidents may go not merely undetected but also unreported.

As an example of this problem, we discuss our ongoing investigation of the economics of reissuing credit cards after a data breach.[15] When a data breach involving credit cards has been announced but the cards have not yet been used for fraud, the banks that issued the compromised credit cards must decide whether to cancel those cards and reissue new ones. Affecting this decision is the fact that the card brands' operating agreements allow issuers to recoup from breached merchants some, but not all, of the issuers' losses from fraud and from reissuing cards.[16]

---

[14] See Alessandro Acquisti, Allan Friedman, and Rahul Telang, *Is There a Cost to Privacy Breaches? An Event Study* *2 (Twenty Seventh International Conference on Information Systems, 2006), archived at http://perma.cc/4QM6-VHT2 (noting that some companies that have experienced "privacy debacles" were subjected to "public outrage and hard to quantify reputation losses").

[15] See generally James T. Graves, Alessandro Acquisti, and Nicolas Christin, *Should Payment Card Issuers Reissue Cards in Response to a Data Breach?* (13th Annual Workshop on the Economics of Information Security, June 2014), archived at http://perma.cc/HJ6E-6Z6V.

[16] See, for example, *Visa International Operating Regulations* *648–53 (Visa, Apr 15, 2014), archived at http://perma.cc/79X7-EDGL. See also Adam J. Levitin, *Private Disordering? Payment Card Fraud Liability Rules*, 5 Brooklyn J Corp, Fin & Comm L 1, 14–15 (2010) (outlining the general rules of liability for unauthorized credit card transactions); Richard A. Epstein and Thomas P. Brown, *Cybersecurity in the Payment Card Industry*, 75 U Chi L Rev 203, 214–16 (2008) (describing the "elaborate systems to detect fraud" and the punishments prescribed for noncompliance).

The costs of canceling and reissuing cards are fairly easy for banks to calculate. They include the expenses of pressing and mailing new cards and the administrative overhead costs of notifying cardholders and responding to customer-service requests. Comments from industry analysts and lawsuits by issuers place the cost of reissuing cards at $3 to $25 per card.[17]

On the other hand, banks can choose not to reissue cards, relying on their fraud detection systems to detect and prevent fraud attempts when they happen. This would be the societally optimal choice if the expected per-card cost of fraud (to all parties) were less than the cost of reissuing a card. Determining the expected cost of fraud is easy if one has access to the issuing banks' databases: one could simply flag those cards that were affected by a data breach and watch for fraud. The system would not be perfect, but issuing banks have grown increasingly sophisticated at noticing patterns of fraud that point to a data breach.[18]

---

[17] See, for example, *Pennsylvania State Employees Credit Union v Fifth Third Bank*, 398 F Supp 2d 317, 322 (MD Pa 2005) (stating that the Pennsylvania State Employees Credit Union canceled and reissued 20,029 cards at a total cost of $98,128, or about $5 per card); Maine Bureau of Financial Institutions, *Maine Data Breach Study* *18–20 (Maine Department of Professional & Financial Regulation, Nov 24, 2008), archived at http://perma.cc/D8U9-CMA7 (finding reissuance costs totaling $1,164,200 across 246,479 reissued cards, or an average cost to issuers of $4.72 per card); Maria Aspan and Clare Baldwin, *Sony Breach Could Cost Card Lenders $300 Mln* (Reuters, Apr 28, 2011), archived at http://perma.cc/W6YE-KCMK (reporting that "[e]ach customer request to replace a credit card would cost lenders about $3 to $5 per card," which includes "the new piece of plastic itself, postage, and various customer service costs"); Chris Churchill, *TJX Reacts to Bank Lawsuit; T.J. Maxx Parent in Filing Says TrustCo Failed to Mitigate Injury from Data Breach*, Times Union B9 (Aug 30, 2008) (citing TrustCo as saying that its costs from the TJX breach, including reissuing four thousand debit cards, were up to $20 per affected account); Mark Jewell, *IDs Are a Steal; Thieves Looking for Credit Numbers Set Their Sights on Big Targets*, Vancouver Columbian E1 (Aug 23, 2004) (reporting that Sovereign Bank reissued eighty-one thousand cards twice at a total cost of about $1 million); Denis Paiste, *Compromised Credit Cards Top 100,000*, NH Union Leader B3 (Jan 31, 2007) (reporting that "[v]arious banks and credit unions have said it costs from $5 to $25 per card reissued"); Anne Ravana, *Banks Start Credit Card Reissue; Breach of Databases Prompts Replacements*, Bangor Daily News A4 (Feb 8, 2007) (quoting a Merrill Bank executive as saying that the cost of replacing seventy-one cards was about $14 per card); Eric G. Stark, *Computer Hackers Are Stealing Bank Card Information, but There Is Protection and Some Banks Have Been Aggressive*, Lancaster Sunday News D1 (July 11, 2004) (reporting that Fulton Bank spent $100,000 to replace twenty thousand cards).

[18] Merchants often learn that they have been breached from their banking partners or payment processors. See, for example, Harriet Pearson, Letter to the Office of the Attorney General (Sept 9, 2014), archived at http://perma.cc/YM3P-QUZ6 (stating that Home Depot "received reports from its banking partners and law enforcement that criminals may have hacked its payment data systems"); Brian Krebs, *Staples: 6-Month*

Alas, most of us do not have access to detailed credit card transaction and fraud data. We could, however, *estimate* the expected cost of fraud to a breached card that is not reissued if we had certain publicly available statistics of high-enough quality. The overall calculation would be simple: divide the number of instances of existing-account credit card fraud (that is, unauthorized charges to credit cards) attributable to data breaches by the number of credit cards exposed by data breaches to get the probability that a breached credit card will be used for fraud, then multiply that probability by the average cost of fraud to obtain the expected cost of fraud.[19]

Of the three statistics needed for that calculation, only the average cost of fraud is reasonably well understood. The DOJ's Bureau of Justice Statistics (BJS) included questions about identity theft[20] in several of its National Crime Victimization Surveys from 2004 through 2012.[21] From 2003 through 2006, the Federal Trade Commission (FTC) also commissioned studies of identity theft victimization.[22] Although the reports are not recent (especially the FTC reports), they are of good quality.

An estimate of the total number of credit cards that are exposed in data breaches and not reissued is more uncertain for three reasons, each fundamentally an issue of missing or incomplete data: First, not all breaches are discovered—a problem of missing data, and one that would be difficult to solve. Second, not all breaches that are discovered are publicly disclosed. Although

---

Breach, 1.16 Million Cards (Krebs on Security, Dec 19, 2014), archived at http://perma.cc/86YF-PUPG (implying that the office-supply store Staples may have been informed of a data breach by banks that noticed suspicious activity); Jeff Goldman, *Data Breach at TripAdvisor's Viator Impacts 1.4 Million Users* (eSecurity Planet, Sept 24, 2014), archived at http://perma.cc/68TP-CXLU (reporting that Viator was notified "by its payment card service provider" of a breach).

19 Here, we are already simplifying. A more precise calculation would treat the likelihood of loss and the cost of loss as probability distributions, which may not be independent.

20 Existing-account credit card fraud, which is the use of stolen credit cards to make unauthorized charges, is classified as a form of identity theft. This is distinguished from new-account fraud, which involves the use of identifying information to open new accounts in the victim's name. See Lynn Langton, *Identity Theft Reported by Households, 2005-2010* *1 (DOJ, Nov 2011), archived at http://perma.cc/4BFK-MKQZ.

21 See, for example, id; *Consumer Sentinel Network Data Book for January – December 2014* *12 (FTC, Feb 2015), archived at http://perma.cc/7B3D-9Y2R; Lynn Langton and Michael Planty, *National Crime Victimization Survey Supplement: Victims of Identity Theft, 2008* *1 (DOJ, Dec 2010), archived at http://perma.cc/3QAP-KDDU.

22 See generally, for example, Synovate, *Federal Trade Commission – 2006 Identity Theft Survey Report* (FTC, Nov 2007), archived at http://perma.cc/BNJ3-PDF2; Synovate, *Federal Trade Commission – Identity Theft Survey Report* (FTC, Sept 2003), archived at http://perma.cc/ZZD3-X5FC.

forty-seven states, the District of Columbia, Guam, Puerto Rico, and the Virgin Islands have laws requiring organizations to notify data subjects when their information has been exposed,[23] these laws have differing reporting triggers and most do not require breaches to be publicized or reported to a public official.[24] This is an incomplete-data problem, but one that might be addressed through legislation. Third, breach notifications often do not disclose how many records were exposed,[25] and those that do may give only rough estimates. In our database of 776 publicly announced credit card data breaches since 2005,[26] only 308—just short of 40 percent—include estimates of the overall number of records affected. Of the 468 breaches in our database in which the breached organizations did not disclose the overall number of records affected, half of the organizations that were breached

---

[23] *Security Breach Notification Laws* (National Conference of State Legislatures, June 11, 2015), archived at http://perma.cc/86QR-KUXN. The three states without data breach notification laws are Alabama, New Mexico, and South Dakota.

[24] Data breach notification laws in California, Connecticut, Florida, Indiana, Iowa, Louisiana, Maine, Maryland, Massachusetts, Missouri, New Hampshire, New York, North Carolina, Vermont, and Virginia require notification of the state attorneys general in the event of a breach that triggers notice requirements. Cal Civ Code § 1798.82(f); Conn Gen Stat Ann § 36a-701b(b)(2)(A); Fla Stat Ann § 501.171(3)(a); Ind Code § 24-4.9-3-1(c); Iowa Code § 715C.2(8); 16 La Admin Code pt III, § 701(A); 10 Me Rev Stat Ann § 1348(5); Md Comm Code Ann § 14-3504(h); Mass Gen Laws Ann ch 93H, § 3(b); Mo Rev Stat § 407.1500.2(8); NH Rev Stat Ann § 359-C:20.I(b); NY Gen Bus Law § 899-aa(8)(a); NC Gen Stat § 75-65(e1); 9 Vt Stat Ann § 2435(b)(3); Va Code § 18.2-186.6(E). Hawaii and South Carolina require notice to be sent to the state departments of consumer affairs. Hawaii Rev Stat § 487N-2(f); SC Code Ann § 39-1-90(K). New Jersey requires notice to the state police. NJ Stat Ann § 56:8-163(c)(1). In all, at least eighteen states require that some state entity be notified in the event of a data breach.

[25] Several states do, however, require breached entities to report the number of the states' residents notified or believed to be affected by a breach. See Fla Stat Ann § 501.171(3)(b)(2); 16 La Admin Code pt III, § 701(A); Mass Gen Laws Ann ch 93H, § 3(b); NH Rev Stat Ann § 359-C:20.I(b); NY Gen Bus Law § 899-aa(8)(a); NC Gen Stat § 75-65(e1); 9 Vt Stat Ann § 2435(b)(3)(C)(i). A few other states' attorney general offices request that information. See generally, for example, *Maine Security Breach Reporting Form*, archived at http://perma.cc/B6DX-62NS; *Guidelines for Businesses to Comply with the Maryland Personal Information Protection Act* (Maryland Attorney General), archived at http://perma.cc/4JGG-HALD; *Database Breach Notification Requirements*, archived at http://perma.cc/FEH3-XTDU (requesting information to be sent to the Virginia attorney general office).

[26] Our database was built by augmenting the Privacy Rights Clearinghouse list of breaches with the breach notifications sent to the attorneys general of Maine, Maryland, and New Hampshire. See *Chronology of Data Breaches: Security Breaches 2005 - Present* (Privacy Rights Clearinghouse), archived at http://perma.cc/G7V9-QPTS; *Data Breach Notices*, archived at http://perma.cc/4L2L-VUGE (listing breach notifications sent to the attorney general of Maine); *Maryland Information Security Breach Notices - 2015* (Maryland Attorney General), archived at http://perma.cc/U8C6-BL7T; *Security Breach Notifications* (New Hampshire Department of Justice), archived at http://perma.cc/PW7M-24JN.

(234) were able to report to the attorneys general of Maine, Maryland, and New Hampshire how many of those states' residents were affected,[27] suggesting that many breached organizations knew the number of people affected by the breaches but chose not to disclose those numbers unless required to do so. This, too, is an incomplete-data problem that might be addressed through legislation.[28]

Because reissued cards do not enter the ecosystem of breached cards that might be used for fraud, the number of cards exposed by data breaches depends on the proportion of those cards that are immediately reissued. That proportion is, like most of the public data about data breach, not well-known. This is more a problem of data access than of data quality: banks know this number, but, considering it proprietary information, they do not share it with the public.[29]

Of the statistics necessary to estimate the total number of breached credit cards that were not reissued—including disclosed breached records, undisclosed breached records, undetected breached records, and the percentage of breached records for which cards were reissued—the first is available only as partial data and the other three are publicly unknown.[30] We can set plausible ranges for these unknowns as parameters in our estimates and we do so in our model, but the lack of data has a significant impact on the quality of our results.

The third statistic used in our estimation is the number of instances of existing-account credit card fraud that are attributable to data breaches. This statistic has two components: the

---

[27] Note, however, that this proportion is biased by the fact that we used the Maine, Maryland, and New Hampshire lists of breach notification letters to extend our database. Because many breach events appeared in one of these lists but not in other sources, those events are more likely not to have total record counts than those that were in our original source database. Of the 477 breaches in the source database after January 1, 2008, when the Maine and New Hampshire reporting took effect, 93 of the organizations (about 24 percent) reported to Maine, Maryland, or New Hampshire the number of residents of those states who were affected but did not disclose an estimate of the overall number of people affected by the breach.

[28] We do not intend to imply that enacting legislation would be easy, especially when a solution to poor data would require either multiple states to modify their data breach notification laws or the federal government to enact a data breach notification law with certain features. But compared to the problem of detecting undetected breaches, legislation has the advantage of being a solution that, while perhaps infeasible, is at least possible. Whether it would be desirable is yet another matter.

[29] We attempted to contact several issuing banks. The two that were willing to talk with us said that their card reissue statistics were confidential information.

[30] For the sake of brevity, we omit some factors that went into the full estimation.

overall scope of credit card fraud, which is well understood, and the proportion of that fraud that is attributable to data breaches, which is not. The scope of credit card fraud is of course the subject of much study. The BJS reports, for example, include estimates of the number of households in which any member was a victim of existing-account credit card fraud.[31]

What is less well-known is the extent to which credit card fraud is attributable to data breaches. This lack of knowledge is not for lack of effort. Several studies, including ones conducted by the FTC and the BJS, have asked victims of identity theft if they knew how their information was obtained. The most common single answer was almost always "I do not know," which accounted for 47 percent to 65 percent of all responses[32] in all but one of the surveys.[33] The percentage of respondents who said that they knew how their information was obtained and that a data breach was the source of that information ranged from 3 percent to 13 percent.[34]

One of the issues with these surveys is what to do about the large number of respondents who have no idea how their information was obtained. Some surveys simply discarded these answers and scaled up the answers of those who said that they knew how their information was obtained.[35] That would be a valid approach if the point of compromise were uncorrelated with the victim's knowledge of that point of compromise, but clearly there is a correlation—a victim of credit card fraud is likely to

---

31 See Langton, *Identity Theft* at *5 (cited in note 20).

32 See Langton and Planty, *National Crime Victimization Survey* at *13 (cited in note 21); Gary R. Gordon, et al, *Identity Fraud Trends and Patterns: Building a Data-Based Foundation for Proactive Enforcement* *3 (Center for Identity Management and Information Protection, Oct 2007), archived at http://perma.cc/6ABG-ZZAH; Synovate, *2006 Identity Theft Survey Report* at *30 (cited in note 22).

33 See Linda Foley, et al, *Identity Theft: The Aftermath 2009* *15 (ITRC, 2010), archived at http://perma.cc/GV2E-F4QV. Seventy-nine percent of respondents to the Identity Theft Resource Center (ITRC) survey said that they knew how their data had been obtained. See Graves, Acquisti, and Christin, *Should Payment Card Issuers Reissue Cards* at *19 (cited in note 15). But ITRC acknowledges that "the fact that ITRC is listed as a victim resource by many entities which have suffered a breach" may "skew" their survey results. Foley, et al, *Identity Theft* at *15 (cited in note 33).

34 Graves, Acquisti, and Christin, *Should Payment Card Issuers Reissue Cards* at *19 (cited in note 15). In the survey by the Center for Identity Management and Information Protection, 53 percent of respondents said they knew how their information was obtained; 50 percent of those said that the source was a business. Id. Because the category "business" includes both breach and nonbreach types of disclosure, we do not include it in the range of results.

35 See, for example, Gordon, et al, *Identity Fraud Trends* at *53 (cited in note 32).

know if his wallet was stolen, for example, but may not know if a skimmer was used.

The large number of parameters in our model, combined with the uncertainty inherent in those parameters because of poor data sources, led to a final estimate with a huge range of possible values. We currently estimate that the expected per-card cost of fraud by not reissuing cards would be somewhere between $0.42 and $310.00 depending on modeling assumptions—a range encompassing three orders of magnitude. According to our model, not reissuing cards would have a social benefit of up to $24 per card or a cost of over $300 per card. When extrapolated over the number of credit cards believed to be exposed in data breaches, this corresponds to an overall savings of $1 billion or an overall cost of $14 billion by not reissuing cards.

Monte Carlo simulations[36] show that our model is particularly sensitive to the cost of existing-account credit card fraud and to the percentage of existing-account credit card fraud attributable to breaches.[37] Given the relative precision with which the financial cost of existing-account credit card fraud is known, the extent to which it is caused by breach is likely the most important unknown factor in determining whether the first-order costs of reissuing cards are more or less than the expected costs of fraud. But this is also one of the least understood of the parameters.

The preceding analysis considers what we might call "first-order costs": the direct costs to cardholders, merchants, and issuers from reissuing cards or from fraud. But indirect costs—which we might term "second-order costs"—are also important. Some indirect costs of not reissuing cards could include increased incentives to engage in credit card breach, more-difficult fraud detection and attribution as a result of stolen credit card data being held longer before use, and reduced credit card use from cardholders who are reluctant to use their cards after a breach.[38] The-

---

[36] For an introduction to Monte Carlo simulations, see W.K. Hastings, *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*, 57 Biometrika 97, 97–98 (1970) ("For numerical problems in a large number of dimensions, Monte Carlo methods are often more efficient than conventional numerical methods.").

[37] See Graves, Acquisti, and Christin, *Should Payment Card Issuers Reissue Cards* at *26–27 (cited in note 15).

[38] See, for example, Shirley W. Inscoe, *Global Consumers React to Rising Fraud: Beware Back of Wallet* *17 (Aite Group, Oct 2012), archived at http://perma.cc/TZ4B-QB2F ("33% of consumers who received replacement cards [after a breach] state that they used

se costs—especially reduced credit card use—may actually have a larger effect than the direct, first-order costs.[39]

B. Inaccurate Data

As an example of the problem of inaccurate data, consider online crime. Online crime has traditionally been thought of as hacking: the unauthorized access to and exploitation of digital property such as electronic credentials. Increasingly, however, online crime consists of traditional criminal activities (such as the sale of counterfeit goods or narcotics) moving online.[40] In fact, since the beginning of the twenty-first century, online criminal activities have become a measurable economic activity.[41]

Criminology and economic scholarship about offline crime has a long history. Often, that scholarship has faced significant data-collection challenges. For instance, to obtain pricing information about narcotics, researchers typically had to rely on undercover agents making drug purchases.[42] The rise of online crime changes that picture—at least in theory. When crimes are committed online, they often leave durable digital trails available for forensic investigation and analysis.[43] In practice, however, the Internet creates its own new data challenges, such as those associated with attribution—the act of determining the party responsible for a security attack. As a result, the belief that estimates of online crime are easier to obtain and more reliable than their offline counterparts may be misplaced.

As an example, Professor Ross Anderson and his colleagues recently attempted to measure the cost of cybercrime, making

---

the new card less frequently than the original card."). It is unclear whether this is because a card was reissued or because of the card exposure regardless of reissue.

39 See Graves, Acquisti, and Christin, *Should Payment Card Issuers Reissue Cards* at *28–30 (cited in note 15).

40 See Tyler Moore, Richard Clayton, and Ross Anderson, *The Economics of Online Crime*, 23 J Econ Persp 3, 3–4 (Summer 2009) (describing the transition from "amateur hackers who defaced websites and wrote malicious software" in a "cottage industry" to "criminal networks" and "online black markets").

41 See id; Ross Anderson, et al, *Measuring the Cost of Cybercrime*, in Rainer Böhme, ed, *The Economics of Information Security and Privacy* 265, 296–97 (Springer 2013).

42 See, for example, Yuehong Yuan and Jonathan P. Caulkins, *The Effect of Variation in High-Level Domestic Drug Enforcement on Variation in Drug Prices*, 32 Socio-Econ Planning Sci 265, 266 (1998).

43 See, for example, Moore, Clayton, and Anderson, 23 J Econ Persp at 8 (cited in note 40) (describing the ability of researchers to "stud[y] the new criminal markets directly" by "monitor[ing] the public chat channels used by online criminals to contact each other," "infiltrat[ing] . . . botnet[s]," and using related means).

explicit all the steps in their measurement methodology.[44] Their estimates ended up being far, far smaller than those provided by professionals in the security industry. One such industry study had put the global cost of cybercrime at $1 trillion—a figure later repeated by the director of the NSA and others, despite the figure's questionable origins,[45] to justify devoting increased resources to deal with online crime. Anderson and his colleagues, while cautioning that "it is entirely misleading to provide totals lest they be quoted out of context,"[46] estimated that the cost of what they call "genuine cybercrime" was in the vicinity of $3.5 billion worldwide at the time of their study.[47] Even considering other, indirect costs (such as loss of consumer confidence) that dwarf the direct costs of cybercrime, their computations appear to remain firmly in the billion-dollar range.

Why the big difference? One obvious possibility is that there are incentives for security professionals that discourage providing conservative estimates of the cost of cybercrime. For instance, declaring that a specific activity will cost society $1 trillion helps make the case for extensively funding countermeasures. And yet the presence of a "price tag," however unreliable it may be, provides a sense of certainty: the argument is seemingly backed by data. In turn, these estimates end up affecting the decisionmaking of other stakeholders—from the share of a company's budget that its CEO allocates for security to national investments in cyberdefense. That cybercrime estimates are often so unreliable may explain why the debate over whether we are spending too little or too much on cyberprotection—which started in the early days of the economics of information security[48]—does not seem to abate.

---

44 Anderson, et al, *Measuring the Cost* at 267–73 (cited in note 41).

45 See Peter Maass and Megha Rajagopalan, *Does Cybercrime Really Cost $1 Trillion?* (Pro Publica, Aug 1, 2012), archived at http://perma.cc/6UD3-GA3S (questioning McAfee's $1 trillion estimate, a figure cited by General Keith Alexander, then-director of the NSA, but published only in a McAfee news release and not in its cybercrime study); Josh Rogin, *NSA Chief: Cybercrime Constitutes the "Greatest Transfer of Wealth in History"* (Foreign Policy, July 9, 2012), archived at http://perma.cc/G2UL-EPW8.

46 Anderson, et al, *Measuring the Cost* at 295 (cited in note 41).

47 Id at 294.

48 Compare Ross Anderson, *Unsettling Parallels between Security and the Environment*, archived at http://perma.cc/A8Z6-L67K ("My intuition is that many firms get it about right, or if anything spend slightly too much [on network security]."), with Bruce Schneier, *Computer Security: It's the Economics, Stupid* (Workshop on Economics and Information Security, May 16, 2002), archived at http://perma.cc/RWR3-J4XQ (arguing that organizations spend too little on computer security but will spend more only if software companies improve their security abilities).

A related example comes from efforts to estimate activity in illicit online markets. A common technique is to estimate the popularity of a given market by looking at the number of listings that are posted in that market.[49] Unfortunately, such listings can generally be created for free, so they are extremely easy to fake. In fact, the operators of a fledgling market have a strong incentive to create or allow fake listings simply to give visitors the impression that significant activity is already taking place. Counting listings has been shown to be an extremely poor predictor of actual transaction volumes,[50] yet the press has repeatedly used this metric as indicative of actual popularity trends.[51]

Poor estimates are not unique to information security. Professors Peter Andreas and Kelly Greenhill give a number of examples in the political science realm and argue that numbers are generally accepted as irrefutable arguments and are thus frequently used in support of policymaking; this in turn creates an "imperative to generate numbers [that] prioritizes bad data over no data."[52] However, this argument applies even more forcefully to computer and information security, because measurements of security incidents may sometimes be simpler in the digital realm—where information is often preserved—than in the physical realm. For instance, a drug purchase on an online

---

[49] See, for example, *Busted, but Not Broken: The State of Silk Road and the Darknet Marketplaces* *1 (Digital Citizens Alliance), archived at http://perma.cc/9PLU-GGHG ("Approximately 13,648 listings for drugs are now available on Silk Road.").

[50] See Kyle Soska and Nicolas Christin, *Measuring the Longitudinal Evolution of the Online Anonymous Marketplace Ecosystem* *38–39 (USENIX, Aug 2015), archived at http://perma.cc/N7V3-WW8M.

[51] See, for example, Patrick Howell O'Neill, *Dark Net Markets Offer More Drugs Than Ever Before* (Daily Dot, May 13, 2015), archived at http://perma.cc/432X-46JT ("Dark Net markets have grown over 37 percent in product listings in the last year despite sweeping police actions and the constant threat of multimillion-dollar thefts looming large."); Steven Nelson, *Largest Online Drug Market Shuts in Massive Suspected Scam* (US News, Mar 18, 2015), archived at http://perma.cc/2MS4-JQLX ("The 15 largest deep web markets had about 42,000 drug listings, indicating the overall deep web drug market had nearly recovered from last year's FBI-led raids."); Jay Newton-Small, *The Silk Road Is Back: The Dread Pirate Roberts Sails the Illicit Online Drug Trade Again* (Time, Apr 30, 2014), archived at http://perma.cc/DX9Q-DZPM ("There are approximately 13,648 listings for drug on the Darknet website . . . compared to the 13,000 listed shortly before [Ross] Ulbricht's arrest."); Matthew Mosk, *Underground Website Used for Black Market Drug Sales Bigger Than the Original, Report Says* (ABC News, Apr 30, 2014), archived at http://perma.cc/K4UU-ESQ2 (noting that the "'darknet' drug economy as a whole contains 75 percent more listings for drugs" than it did before the FBI shut down the popular site Silk Road).

[52] Peter Andreas and Kelly M. Greenhill, *Introduction: The Politics of Numbers*, in Peter Andreas and Kelly M. Greenhill, eds, *Sex, Drugs, and Body Counts: The Politics of Numbers in Global Crime and Conflict* 1, 19 (Cornell 2010).

anonymous marketplace will usually leave a digital trail that can be used for analysis after the fact.[53] On the other hand, a similar transaction in the physical world would have to be inferred from other signals or directly witnessed at the time it took place.

C. Invalid Inferences

Online blacklists offer one example of the perils one might encounter when drawing inferences from online criminal data. Blacklists are frequently used to flag computers involved in nefarious activities on the Internet. A common use of blacklists is to determine which computers are involved in sending spam e-mail.[54] The idea is that a mail client can check any incoming piece of e-mail against a blacklist. If the computers involved in sending the e-mail are in the blacklist, the e-mail client can then refrain from displaying the suspected spam e-mail, in turn relieving the e-mail user from performing this filtering operation herself.[55]

Andreas Pitsillidis and his colleagues examined several "spam feeds" used for blacklisting and reached the conclusion that different spam feeds performed in very different ways.[56] They show, for instance, that the pairwise intersection of Internet domains listed in two different spam feeds is generally less than 66 percent and can be as low as 1 percent—that is, given two different spam feeds, only 1 percent to 66 percent of the domains listed in either feed would appear in both.[57] As a result, any analysis that extrapolated estimates based on a single spam feed would likely be erroneous. Hence, estimates of spam volume, evolution of spam over time, and other metrics are considerably more challenging than initially envisioned. In turn, this would have repercussions on intervention policies.

For an extreme example, assume that a given blacklist shows that a majority of spam is coming from a certain country—call it Freedonia. As a result, one might push for filtering e-mails coming from Freedonia much more aggressively than

---

[53] See, for example, Andy Greenberg, *Follow the Bitcoins: How We Got Busted Buying Drugs on Silk Road's Black Market* (Forbes, Sept 5, 2013), archived at http://perma.cc/TWY2-VU29.

[54] See *How Blacklists Work* (MailChimp, July 24, 2015), archived at http://perma.cc/534J-MHHL.

[55] See id.

[56] See Andreas Pitsillidis, et al, *Taster's Choice: A Comparative Analysis of Spam Feeds* *1 (ACM, Nov 2012), archived at http://perma.cc/5FCJ-9XA3.

[57] See id at *7–8.

those from other countries. But the blacklist operator may just have better sensors in Freedonia than in other countries. The country of Sylvania, for example, might generate much more spam than Freedonia, but the blacklist operator would not see any of it because the operator simply does not monitor Sylvanian traffic.

For another example of inaccurate data analysis, consider the US government's case against Ross Ulbricht, whom the government accused of running the infamous online bazaar Silk Road.[58] The FBI's original September 2013 criminal complaint included a statement that the total revenue of the site in Bitcoins—the online currency in which Silk Road did all of its business[59]—was equivalent to about $1.2 billion.[60] Although this figure was completely at odds with previous estimates that were much more modest,[61] the FBI's estimate was widely echoed in the media.[62] In January 2015, when the trial began, prosecutors sharply reduced the estimate of Silk Road's total business to about $200 million[63]—a number far more in line with independent estimates carried out by academic researchers.[64]

The sixfold discrepancy came not from incorrect data but from incorrect extrapolations. Bitcoins suffered from an extremely volatile exchange rate with US dollars during Silk

---

58 See Newton-Small, *The Silk Road Is Back* (cited in note 51).

59 Rainer Böhme, et al, *Bitcoin: Economics, Technology, and Governance*, 29 J Econ Persp 213, 222–23 (Spring 2015).

60 Sealed Complaint, *United States v Ulbricht*, Criminal Action No 13-2328, *15 (SDNY filed Sept 27, 2013) ("*Ulbricht* Complaint") ("The total revenue generated from these sales was 9,519,664 Bitcoins, and the total commissions collected by Silk Road from the sales amounted to 614,305 Bitcoins. These figures are equivalent to roughly $1.2 billion in revenue and $79.8 million in commissions.").

61 See, for example, Nicolas Christin, *Traveling the Silk Road: A Measurement Analysis of a Large Anonymous Online Marketplace* *10 (International World Wide Web Conference Committee, May 2013), archived at http://perma.cc/55QM-4MKJ (estimating Silk Road's monthly sales volume at approximately $1.2 million).

62 See, for example, Joshuah Bearman, *Silk Road: The Untold Story* (Wired, May 23, 2015), archived at http://perma.cc/JL8R-3F43; David Segal, *Eagle Scout. Idealist. Drug Trafficker?* (NY Times, Jan 18, 2014), archived at http://perma.cc/FG78-NJR8; Saumya Vaishampayan, *Silk Road Drug Market Handled $1.2 Billion of Transactions in 2.5 Years before FBI Seizure* (MarketWatch, Oct 2, 2013), archived at http://perma.cc/E4T2-JJ46.

63 Emily Flitter, *U.S. Sharply Reduces Silk Road's Estimated Sales Volume* (Reuters, Jan 20, 2015), archived at http://perma.cc/H5N7-P2AV.

64 See, for example, Soska and Christin, *Measuring the Longitudinal Evolution* at *40 (cited in note 50) (estimating Silk Road's gross sales at over $100 million per year, which "appears consistent with the (revised) US Government calculations of $214M of total grossed income by Silk Road over its lifetime").

Road's life span.[65] Because the FBI seized the server that had hosted the marketplace, it had access to actual sales records.[66] Combined with historical information about Bitcoin-to-dollar conversion rates, this data would have allowed the FBI to calculate a total transaction value that was as close as one could get to its actual value. But instead of taking these fluctuations into account, the FBI simply multiplied the number of Bitcoins found on the server by the exchange rate at the time of the complaint.[67] Despite the FBI's access to accurate data, a simplistic methodology grossly overstated the amount of business done on Silk Road.

Ultimately, the miscalculation probably made little difference to Ulbricht's case. Ulbricht was charged not only with running Silk Road but also, in a separate case in Maryland, with the attempted murders for hire of several individuals who threatened to expose the operation.[68] This, plus the court's findings that Ulbricht had the resources to flee and was in possession of false identification documents, was enough to prevent Ulbricht's release on bail.[69] The figure was corrected early in trial and it is unclear whether jurors ever heard the inflated figure. The difference between $200 million and $1.2 billion might have affected sentencing under the charges of money-laundering had it gone uncontested,[70] but even with the lower $200 million estimate, Ulbricht received a sentence of life in prison.[71]

Although the incorrect $1.2 billion estimate may not have mattered to Ulbricht's sentence, the error could nevertheless prove not to be harmless. Consider this example of how incorrect estimates can propagate: the $1.2 billion number already

---

65 See *Ulbricht* Complaint at *15 (cited in note 60).

66 See id at *7 (indicating that the FBI gained access to the Silk Road servers).

67 See id at *15.

68 See *United States v Ulbricht*, 31 F Supp 3d 540, 550 (SDNY 2014) ("According to the Indictment, the defendant pursued violent means, including soliciting the murder-for-hire of several individuals he believed posed a threat to that enterprise.") (quotation marks omitted). See also Superseding Indictment, *United States v Ulbricht*, Criminal Action No 13-0222, *6–11 (D Md filed Oct 1, 2013).

69 Andy Greenberg, *Alleged Silk Road Boss Ross Ulbricht Now Accused of Six Murders-for-Hire, Denied Bail* (Forbes, Nov 21, 2013), archived at http://perma.cc/XCR9-3ADT.

70 The higher number might have put Ulbricht into the highest dollar-value category under the United States Sentencing Guidelines, adding two points to the offense level (such that the dollar value would add thirty points rather than twenty-eight). See United States Sentencing Commission, *Guidelines Manual* §§ 2B1.1(b)(1), 2S1.1(a)(2) (2014).

71 Andy Greenberg, *Silk Road Creator Ross Ulbricht Sentenced to Life in Prison* (Wired, May 29, 2015), archived at http://perma.cc/GAF2-LPNM.

appears in the 2014 UN World Drug Report[72]—which, in turn, is the source of numbers that are widely cited in discussions of drug policy.

The examples above show that even when properly collected, data can be misinterpreted, propagated, and used to derive numbers that are incorrect or policies that are ineffective. Unfortunately, the veneer of scientific soundness coming from proper data collection makes it harder to dispel subsequent data misinterpretations.

II. DISCUSSION

The previous Part highlights some of the difficulties of obtaining and using accurate measurement data in the context of online criminality. Understanding the true costs or benefits of a security decision may depend on estimating parameters whose distributions are not well-known. Some of these unknowns are more unknown than others. In particular, we have considered a series of major types of unknowns, or data problems:

- Data may not exist. In the credit card example, the extent of undetected breach is an example of this type of unavailable data. Or, similarly, data may exist but be inaccessible. For example, we were unable to find estimates of how effective fraud-monitoring software is at preventing credit card fraud. That information is likely well-known by card issuers. The percentage of breached cards that are reissued is another parameter of this type: each issuer knows how many cards it reissues.
- Data may exist but be of poor quality or vary across organizations. The uncertainty surrounding estimates of the extent to which data breaches are a cause of existing-account credit card fraud is an example of this. Several surveys exist, but their results do not shed much light on the actual extent to which data breach is a cause of existing-account credit card fraud.

---

[72] *World Drug Report 2014* *18 (UN Office on Drugs and Crime, June 2014), archived at http://perma.cc/4GKG-4TW8.

- Data may exist but be used incorrectly—for instance, because of biased inferences, inaccurate calculations, or unwarranted generalizations.

Our examples also outline a couple of major possible causes for these measurement difficulties. First, incentives may be misaligned; for instance, vendors of security products have incentives against producing conservative estimates.[73] Second, deriving precise measurements is generally a complex task that may be undermined by collection biases (as in our blacklist example[74]), insidious errors (as in our anonymous black market example[75]), or high sensitivity to hard-to-measure variables (as in the credit card example[76]).

Our examples suggest that policy decisions need to account for the inherent limitations of these measurements instead of blindly relying on sometimes-approximate estimates. Our sensitivity analysis suggests that resources could usefully be targeted toward getting better data for parameters that are critical to our model. Specifically, it would be useful to get better information on how identity thieves get access to credit card data. Surveys of victims are clearly inadequate; too many people simply do not know how their credit card information was obtained. Issuers, however, have the ability to connect breach notification with card misuse. Issuers also have information, at least collectively, on the percentage of cards that they reissue after a breach. Access to these sources of data would undoubtedly improve our understanding of the benefits of the options available following a data breach. Similarly, when issues of data quality occur, sometimes the problems can be solved with greater cooperation from those who hold the data.

Sharing of information regarding attacks, breaches, or vulnerabilities among private sector organizations, as well as between the public and the private sectors, is therefore vital. Unsurprisingly, data sharing has repeatedly been advocated in reports concerning the vulnerabilities of digital infrastructure.[77]

---

73 See Part I.B.

74 See Part I.C.

75 See Part I.B.

76 See Part I.A.

77 See, for example, Fred Chong, et al, *National Cyber Leap Year Summit 2009: Co-Chairs' Report* *9–12 (Sept 16, 2009), archived at http://perma.cc/4CAK-5TZ6 ("[I]t is imperative that appropriate models of cooperation are developed immediately to incentivize

However, our analysis also suggests that although more information might be a necessary condition for improving cybersecurity, it is not sufficient. Having *more* information does not help if the information is the product of inaccurate estimates[78] or if the information is then used for unwarranted generalizations. Rather than the mere sharing of data, our analysis also calls for the sharing of the methodologies, assumptions, and procedures used in the collection, aggregation, and analysis of cybercrime data. An increasing number of journals in the empirical social sciences have asked their authors to do something quite similar, encouraging researchers to make the data and codes used in their studies available to the public.[79]

All of the examples we have discussed pertain primarily to information security. As noted in Part I.B, however, these informational problems are not exclusive to the domain of cybercrime. For instance, perverse incentives may affect a car vendor's promptness in disclosing a newly found potential vehicle-safety issue. But the problems of asymmetric and incomplete information seem nearly endemic in the cybersecurity realm, in which attackers try to hide their traces and attribution of attacks is often hard to establish.

These information problems extend into the realm of national security. Although online crime is primarily the product of relatively complex networks of financially motivated criminals, whereas national security (reportedly) involves nation-state actors that are more sophisticated, large-scale criminal activities do have the potential to impact national security. National-security decisions can also be affected by the problems of imperfect information that we have highlighted above.[80]

---

the participants to engage in research, development, and testing of technologies and approaches to achieve [security] goals.").

78 Consider Professors Andreas and Greenhill's "imperative" to generate numbers, prioritizing bad data over no data. See text accompanying note 52.

79 See, for example, Shirley S. Wang, *How Much Should Scientists Check Other Scientists' Work?* (Wall St J, Oct 5, 2015), archived at http://perma.cc/WN46-JWGT (discussing one psychology journal's initiative to promote data sharing, as well as the advantages and disadvantages of "open science and data sharing").

80 One key difference does exist between online criminality and national security. When dealing with monetary losses, insurance may be available to shift the risk of loss associated with a successful attack. On the other hand, national-security issues may be much harder to insure against. For instance, state secrets cannot be insured—when names and addresses of undercover operatives are revealed to an adversary, they cannot be "unrevealed." Likewise, resources at a national scale, such as an entire cellular network, may simply be too onerous to even consider insuring. This, however, does not fun-

Consider for instance, the recent data breach suffered by the Office of Personnel Management (OPM).[81] The White House has suggested that it could retaliate against a specific nation-state for this attack,[82] which indicates that it has reasonable confidence regarding to whom the attack should be attributed. However, public hearings have uncovered serious security vulnerabilities that were, from a purely technical standpoint, easy enough to exploit that they could have been used by criminals for monetary gain. Along the same lines, our discussion of the Silk Road bazaar in Part I.C should be put into a larger context. Several marketplaces similar to Silk Road have sprung up in its stead,[83] leading to considerable international police cooperation to attempt to shut them down.[84]

These examples show that the line between large-scale criminal activities and national security is often blurry. Consider the Convention on Cybercrime,[85] adopted by the Council of Europe and several other nations (including the United States and Japan).[86] Although the Convention primarily focuses on criminal activities, it also encompasses attacks on critical infrastructure such as dams or power plants—things that are typically considered to be national-security matters.[87] As a result, we expect much of the argument we developed in this Essay in the context of online crime to carry over to the context of national security.

---

damentally change our argument; if anything, the absence of a potential remedy (insurance) further reinforces the need for sound risk-assessment methodologies.

81 For details about the OPM breach, see Andrea Peterson, *OPM Says 5.6 Million Fingerprints Stolen in Cyberattack, Five Times as Many as Previously Thought* (Wash Post, Sept 23, 2015), archived at http://perma.cc/EPZ4-XTVB.

82 See Kellan Howell, *U.S. Will Retaliate against China for OPM Hacks* (Wash Times, Aug 1, 2015), archived at http://perma.cc/TEN9-9FUV.

83 See Soska and Christin, *Measuring the Longitudinal Evolution* at *33–34 (cited in note 50).

84 See *Dozens of Online "Dark Markets" Seized pursuant to Forfeiture Complaint Filed in Manhattan Federal Court in Conjunction with the Arrest of the Operator of Silk Road 2.0* (DOJ, Nov 7, 2014), archived at http://perma.cc/6VCE-GWCR (noting that the seizures were "part of a coordinated international law enforcement action").

85 TIAS No 13174, 2296 UNTS 167 (2001).

86 *Chart of Signatures and Ratifications of Treaty 185: Convention on Cybercrime* (Council of Europe, Oct 19, 2015), archived at http://perma.cc/J72Y-3L6E.

87 See Nicolas Christin, *On Critical Infrastructure Protection and International Agreements* *4–5 (Center for International and Security Studies at Maryland, Mar 2011), archived at http://perma.cc/H7Q3-WXQ2.