

Empirical Analyses of Effects and Perceptions of Cybercrime

Submitted in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in the
Department of Engineering and Public Policy

James T. Graves

B.S., Mathematics/Computer Science, Carnegie Mellon University
M.S., Information Networking, Carnegie Mellon University
J.D., William Mitchell College of Law
LL.M., Georgetown University Law Center

Carnegie Mellon University
Pittsburgh, PA

December 2021

© James T. Graves, 2021
All Rights Reserved

Acknowledgements

I would like to express my deepest appreciation to my committee: Alessandro Acquisti (chair), Nicolas Christin, Lorrie Cranor, and David Thaw. Alessandro has been the kindest and most thoughtful advisor a student could ask for; even when we had disagreements, he was always so darn *nice* about them. I am indebted to my committee for their patience, to Alessandro for nudging me to complete my dissertation, and to David for encouraging me when I had all but decided that I did not really need a Ph.D., anyway.

I have had the honor of working with many other incredible people. Ross Anderson's work on the economics of information security is part of what prompted me to try to join the field myself; getting to co-author a paper with him was like unlocking a security-nerd achievement. Kirsten Marten was a tremendous resource on doing factorial vignette surveys; whatever I managed to do right in Section 3.4 is thanks to her (any mistakes are my own). Alfred Blumstein taught me about criminology and gave me candid feedback on whether what I was originally proposing to do in Chapter 4 was even feasible.

I have been fortunate to be part of the Privacy Economics and Experiments (PeeX) Lab at CMU, and I am constantly amazed at the work everyone in that group is doing. Special thanks to Sasha Romanosky, who always had time to give me great advice early in my academic career; if the title of Chapter 2 seems eerily similar to that of one of Sasha's papers, that is because I saw his work as the standard to aspire to. Thanks also to Alyssa Au, who did a lot of the early data collection of court records from PACER for Chapter 4 and helped with research on CFAA sentencing.

Many thanks to the participants of the 2015 Privacy Law Scholars Conference and 2020 Cybersecurity Law and Policy Scholars Conference for their comments on versions of what would eventually become Chapter 4. I am especially indebted to the commenters on my papers at both conferences: Jonathan Mayer provided tremendously helpful and detailed comments on an preliminary version of that research; Aaron Cooper provided great feedback on a much later version.

Finally, but most importantly, I am extremely grateful to my wife, Christy. Without her support (in many senses of the word), none of this would have been possible.

This work was partially funded by NSF IGERT grant DGE-0903659. Portions of this work were also 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. Latter stages of this work were completed while I was employed at the Electronic Privacy Information Center, the Institute for Public Representation at Georgetown Law, and the Federal Communications Commission. This work does not necessarily represent the positions of any of the aforementioned agencies or organizations.

Abstract

This work investigates three related topics involving some of the consequences or perceptions of cybercrime. These consequences and perceptions have important policy implications, ranging from potentially billions of dollars in societal costs that depend on credit card reissue decisions to the perceived fairness of sentences imposed on people convicted of cybercrimes. I analyze these consequences and perceptions using a combination of legal, empirical economic, and criminological analyses.

First, this work analyzes at the economics of credit card reissue after a data breach. Using a parameterized estimation model based on publicly-available information, it compares the cost of reissuing cards to the total expected cost of fraud if cards are not reissued. The model suggests that automatically reissuing cards may have lower social costs than the costs of waiting until fraud is attempted, although the range of results is considerably broad. The results also show how a lack of quality public information about data breach and identity theft can make informed public policy decisions more difficult.

Second, it explores a potential misalignment between the factors that contribute to cybercrime sentences and the importance of those factors in public perceptions. It presents the results of two empirical studies that measure public perceptions of different factual attributes of cybercrime. The studies show that Computer Fraud and Abuse Act (CFAA) sentences are indeed out of alignment with the public's views and provide empirical support for arguments that CFAA sentencing is miscategorized in the federal sentencing guidelines.

Third, it addresses the question of whether CFAA subsection (a)(2), covering unauthorized access to a computer to obtain information, should be considered a trespass, burglary, or fraud statute. It does this through an analysis of case records and sentencing data from 1,095 real-world CFAA sentences, and an experimental study of perceptions of 499 participants on Amazon Mechanical Turk. The results of both studies suggest that (a)(2) is not like trespass or fraud, at least in terms of current punishments or perceptions, and lend support to arguments that CFAA sentencing should be covered under its own section of the sentencing guidelines.

Table of Contents

1. Introduction	1
2. Should Credit Card Issuers Reissue Cards in Response to a Data Breach? Uncertainty and Transparency in Metrics for Data Security Policymaking	4
2.1. Introduction	4
2.2. Background	5
2.3. Methodology	9
2.4. Data	12
2.4.1. The Cost of Reissuing Cards.....	13
2.4.2. The Probability of Credit Card Misuse Following a Breach	13
2.4.3. The Cost of Credit Card Fraud.....	20
2.5. Analysis.....	21
2.5.1. Monte Carlo Analysis	22
2.5.2. Sensitivity Analysis.....	23
2.6. Discussion and Limitations	24
2.6.1. Limitations and Opportunities for Future Work	25
2.6.2. Conclusion	26
3. Perception Versus Punishment in Cybercrime.....	27
3.1. Introduction	27
3.2. Background	29
3.2.1. Factors Affecting Sentencing Under the Computer Fraud and Abuse Act.....	29
3.2.2. Criminological Studies of Crime Seriousness	34
3.3. Study I: Between-Subjects Experiments.....	37
3.3.1. Methodology	38
3.3.2. Theoretical Model	41
3.3.3. Results	41
3.4. Study II: Factorial Vignette Survey Experiment.....	47
3.4.1. Methodology	48
3.4.2. Theoretical Model	51
3.4.3. Results	51
3.5. Discussion	53
3.5.1. Comparison of Results Between the Two Studies	53
3.5.2. Implications for Sentencing Policy	54
3.5.3. Limitations and Opportunities for Further Research	59
3.5.4. Conclusion	59
4. An Empirical Analysis of Sentencing and Perceptions of “Access to Information” Computer Crimes	61
4.1. Introduction	61
4.2. Background and Related Work	62
4.2.1. Computer Crime Norms.....	62
4.2.2. Empirical Analyses of Cybercrime	66
4.3. Analysis of 1030(a)(2), Trespass, Burglary, and Fraud Sentences	68
4.3.1. Data and Methodology	68
4.3.2. What do 1030(a)(2) Computer Crimes Look Like?	71
4.3.3. What do Federal Trespass Crimes Look Like?	83

4.3.4. What do Federal Burglary Crimes Look Like?.....	88
4.3.5. What do Federal Fraud Crimes Look Like?.....	89
4.3.6. CFAA Compared to Burglary, Trespass, and Fraud.....	92
4.4. Public Perceptions.....	96
4.4.1. Methodology.....	97
4.4.2. Results.....	98
4.5. Discussion.....	102
4.5.1. Policy Implications.....	102
4.5.2. Limitations and Opportunities for Future Work.....	102
4.5.3. Conclusion.....	104
5. Conclusion.....	106
Appendix A. U.S. Sentencing Guidelines Sentencing Table.....	108
Appendix B. Regression Tables for the Chapter 3 Between-Subjects Experiments.....	109
Appendix C. Inter-Respondent Heterogeneity in the Chapter 3 Factorial Experiments.....	115
Appendix D. Example Survey Text for the Chapter 3 Between-Subjects Experiments.....	118
Appendix E. Survey Text for the Chapter 3 Factorial Vignette Survey.....	128
Appendix F. Distribution of Ratings by Vignette in the Chapter 4 Perceptions Study.....	135
Appendix G. Data Quality Checks for the Chapter 4 Perceptions Study.....	138
G.1. Inconsistently Rated Vignette Pairs.....	138
G.2. Overall Distribution of Ratings.....	141
G.3. Effect of Completion Time on Results.....	142
G.4. Ordering Effects.....	143
G.5. Correlation Between Rating Types.....	145
G.6. Regression Results When Excluding “Unreliable” Responses.....	146
G.7. Regressions by Individual Vignettes.....	148
Appendix H. Survey Text for the Chapter 4 Perceptions Study.....	150

Tables

Table 1: Estimated total number of credit card records exposed in data breach per year.....	17
Table 2: Calculation of the probability of existing-account credit card fraud to an account affected by a breach.....	20
Table 3: Expected cost per card of an existing-account credit card fraud incident	21
Table 4: Comparison of the per-card cost of reissuing vs. not reissuing cards.....	22
Table 5: CFAA Sections and Maximum Sentences	30
Table 6: Summary of Regression Results in Between-Subjects Experiments	43
Table 7: Pairwise correlation matrix for the DVs in the between-subjects experiments	47
Table 8: Mixed-effects regression for the factorial experiment.....	52
Table 9: Impact of offense factors on perceptions and sentences.....	55
Table 10: Sentencing examples for the factorial scenario	57
Table 11: CFAA convictions by subsection, 1/31/05– 12/31/18	75
Table 12: Number of 1030(a)(2) sentences by access type and year	78
Table 13: CFAA Sentencing by subsection	79
Table 14: 1030(a)(2) sentencing by access type	79
Table 15: 1030(a)(2) sentencing by fact pattern	80
Table 16: Co-statutes of conviction with 1030(a)(2)	81
Table 17: Application of the “Special Skill” and “Sophisticated Means” enhancements by access type for (a)(2) sentences.....	82
Table 18: Application of the “Special Skill” and “Sophisticated Means” enhancements by fact pattern for (a)(2) sentences.....	83
Table 19: Sentencing Guideline Factors for Fraud, Burglary, and Trespass Crimes.....	92
Table 20: Demographics by crime type (single count of conviction).....	93
Table 21: Sentencing by crime type (single count of conviction)	94
Table 22: Regression results comparing CFAA and non-CFAA fraud sentences	95
Table 23: Departures from the sentencing guideline ranges, by crime type.....	96
Table 24: CFAA vignettes	97
Table 25: Trespass vignettes	97
Table 26: Burglary vignettes.....	97
Table 27: Fraud vignettes.....	98
Table 28: Regression results for perceptions of crime types	101
Table 29: U.S. Sentencing Guidelines Sentencing Table	108
Table 30: Ordered probit marginal effects for the Type of Data experiment	109
Table 31: Ordered probit regression results for the Scope experiment.....	110
Table 32: Ordered probit regression results for the Motivation experiment (vs. Profiteer)	111
Table 33: Ordered probit regression results for the Consequences experiment (vs. Low).....	112
Table 34: Ordered probit regressions for the Co-Responsibility experiment	113
Table 35: Ordered probit regressions for the Context experiment (vs. Bank).....	114
Table 36: Statistically-significant coefficients as percentages of individual-level regressions....	116
Table 37: Summary statistics for coefficients across individual-level regressions.....	116
Table 38: Summary of data-quality exclusion criteria.....	147
Table 39: Regressions with exclusions	147
Table 40: Regressions by individual vignette, with “loose” exclusions	148
Table 41: Regressions by individual vignette, with “strict” exclusions	149

Figures

Figure 1: Credit card payment network structure	5
Figure 2: Survey and study results for the number of identity theft victims knew how their data was obtained, and, if so, the point of compromise	19
Figure 3: Distribution of the cost per card to reissue or not reissue cards based on a Monte Carlo simulation	22
Figure 4: Histogram of cumulative savings from not automatically reissuing cards according to a Monte Carlo simulation	23
Figure 5: Tornado diagram of variables affecting the per-card cost of not reissuing cards	24
Figure 6: Factorial instrument rating task slider	50
Figure 7: CFAA (a)(2) fact patterns in the PACER data set	71
Figure 8: Number of CFAA sentences by year, 1/31/05– 12/31/18	77
Figure 9: Summary of vignette responses	100
Figure 10: Distribution of β values over respondent-level models for non-log-scaled variables of interest	117
Figure 11: Distribution of β values over respondent-level models for log-scaled variables of interest	117
Figure 12: Distribution of ratings by vignette: Seriousness	135
Figure 13: Distribution of ratings by vignette: Wrongfulness	136
Figure 14: Distribution of ratings by vignette: Harmfulness	137
Figure 15: Distribution of IRDVPs per Respondent	139
Figure 16: Distribution of IRDVPs where the difference in ratings is more than 5 points on the 100-point scale	139
Figure 17: Inconsistently rated IRDVPs as a percent of all DVPs, by ordering distance	140
Figure 18: Overall distribution of ratings	141
Figure 19: Distribution of “100” ratings per respondent	141
Figure 20: IRDVPs (by 10 or more) by how long the respondent took to answer the survey	142
Figure 21: Number of “100” Ratings by Completion Time	143
Figure 22: Distribution of Responses by Display Order	143
Figure 23: P values of individual regressions on ratings types (DV: Seriousness)	145
Figure 24: Distribution of R^2 values from individual regressions on rating types	146

1. Introduction

For a period of about four years, Albert Gonzalez was the most prolific credit card thief in the world. From 2003 to 2006, as the ringleader of a conspiracy that hacked into TJX, Heartland Payment Systems, Office Max, and others, Gonzalez and his confederates stole over 100 million credit and debit card numbers. The TJX heist alone cost the retailer over \$170 million, and credit card issuers alleged costs of up to \$25 per card from reissuing cards—a total of \$2.5 billion if all 100 million cards were reissued.¹ Gonzalez himself made “well over \$1 million from the scheme.”² Gonzalez was eventually arrested and, after a plea bargain, sentenced to 20 years in prison.³

One of Gonzalez’s confederates was his friend Stephen Watt, who adapted a sniffer program for use in the TJX hack. Watt received no money for his part in the crime. Instead, as the court responsible for sentencing him explained, Watt “did it for the challenge, for the thrill of besting large institutions.”

Watt was a first-time offender involved in a conspiracy that led to losses in the hundreds of millions, if not billions, of dollars. That Watt’s first crime was a big one led to vastly different sentencing recommendations. Watt’s attorneys asked for a sentence of six months probation. Government attorneys requested the statutory maximum sentence of five years in prison. Had the statute not set that maximum sentence, the sentencing guidelines would have called for a life sentence because of the amount of loss. The court sentenced Watt to two years in prison with three years of supervised release and ordered restitution of \$171.5 million.⁴

Watt’s case illustrates some of the questions that arise in public policy regarding cybercrime.⁵ Some of those questions are inherent to any crime, such as the roles of offender culpability, motivation, harm, and fairness in sentencing. Others seem unique to the cybercrime context, or at least more pronounced. For example, would it really have cost issuers \$2.5 billion to reissue all the cards that were breached? If so, would the cost of fraud on those cards actually be less than that? How large of a role should the outsized losses play in a cybercrime, in which such losses are easier to create? Should the lack of physical danger mitigate against harsh sentences? Is the sophistication of nearly all cybercrimes reason to increase sentences? Should it matter that TJX arguably shared some responsibility for the crime by not adequately securing its network? Many of these questions are normative, but they can be investigated using empirical methods.

Although a robust field of research has grown around the economics of data security—and with it, cybercrime—attitudes towards cybercrime have been relatively less explored. And

¹ Amended Consolidated Class Action Complaint, *In re TJX Companies Retail Security Breach Litigation*, 527 F. Supp. 2d 209 (D. Mass. 2007) (No. 07-10162).

² *United States v. Watt*, 707 F. Supp. 2d 149, 154 (D. Mass. 2010).

³ *Id.* at 150 n.2.

⁴ *Id.* at 151.

⁵ In this thesis, I use the terms “cybercrime” and “computer crime” interchangeably.

despite a rich literature investigating perceived seriousness of a broad set of crimes,⁶ the seriousness of cybercrimes has received less study. Furthermore, although legal scholars have debated the how to analogize cybercrimes to real-world crimes, there appears to have been no empirical or experimental investigations into a question that turns out to be difficult to answer: just what *is* cybercrime, anyway? Is it a fraud crime, as it is treated for purposes of sentencing? Is it more like trespass, as it is frequently compared to? Or is it something else entirely?

The answers to these questions can have important public policy implications. With respect to reissuing credit cards, billions of dollars in social costs may hang on issuers' decisions, which in turn depend on incentives built into private ordering in the form of card association agreements. Furthermore, the problem of estimating the social costs highlights the need for better data with which to make data security policy decisions. Attitudes about cybercrime can inform different policy decisions about which attributes of a cybercrime should be used to enhance sentences for offenders. For example, evidence that for-profit cybercrimes are perceived to be more serious than cybercrimes motivated by activism would provide support for sentencing enhancements based on motive. And how cybercrimes are punished and perceived compared to real-world crimes goes to the heart of fairness in sentencing.

This thesis investigates three related topics involving some of the consequences or perceptions of cybercrime.

Chapter 2 looks at the economics of credit card reissue after a data breach. Issuers spend millions of dollars each year reissuing credit cards that were exposed in data breaches. But are the social costs of reissuing these cards lower than the expected costs of fraud if cards are not reissued? Liability allocation rules enforced by the credit card brands may reimburse fraud losses but not the cost of reissuing cards. If reissuing incurs lower social costs, those allocation rules could be inefficient. Although issuers may evaluate the internal risks and benefits of reissuing, to my knowledge this work represents the first attempt to measure the merits of each option when costs external to the issuers are considered. Using a parameterized model and Monte Carlo simulation, Chapter 2 compares the cost of reissuing cards to the total expected cost of fraud if cards are not reissued. The ranges and distributions in the model are informed by publicly-available information, from which I extrapolate estimates of the number of credit card records historically exposed in data breaches, the probability that a card exposed in a breach will be used for fraud, and the associated expected cost of existing-account credit card fraud. The model suggests that automatically reissuing cards may have lower social costs than the costs of waiting until fraud is attempted, although the range of results is considerably broad.

Chapter 3 explores a potential misalignment between the factors that contribute to cybercrime sentences and the importance of those factors in public perceptions. How society—including victims and potential victims of cybercrime—views cybercrime can affect how cybercrimes are defined, what punishments they carry, whether those punishments are believed

⁶ See *infra* Section 3.2.2,

to be fair, and how resources are allocated to enforcement.⁷ They can also inform the important question of whether crimes under the Computer Fraud and Abuse Act (CFAA) should be sentenced as fraud crimes, as they are now, or whether a different standard should apply. Chapter 3 reports on the results of two empirical studies that measure public perceptions of different factual attributes of cybercrime. The studies show that CFAA sentences are indeed out of alignment with the public's views. In particular, the amount of loss attributed to a CFAA crime plays a much larger role in the sentencing guidelines than is reflected in public attitudes, while the attacker's motivation—which is important to public perceptions—plays almost no role in sentencing. These results provide empirical support for arguments that CFAA sentencing is miscategorized in the federal sentencing guidelines.

Chapter 4 addresses the question of whether CFAA subsection (a)(2), covering unauthorized access to a computer to obtain information, is perceived and punished like a trespass, burglary, or fraud statute. It does this through two studies: (1) an analysis of case records and sentencing data from 1,095 real-world CFAA sentences, and (2) an experimental study of perceptions of 499 participants on Amazon Mechanical Turk. The results of both studies suggest that (a)(2) is not like trespass or fraud, at least in terms of current punishments or perceptions. CFAA (a)(2) crimes receive lower punishments and are perceived to be less serious crimes than fraud or burglary crimes. But (a)(2) crimes receive harsher punishments and are perceived to be more serious than federal trespass crimes that do not involve weapons. These results lend support to arguments that CFAA sentencing should be covered under its own section of the sentencing guidelines, at least for (a)(2) offenses.

Chapter 5 concludes.

⁷ Michael O'Connell & Anthony Whelan, *Taking Wrongs Seriously: Public Perceptions of Crime Seriousness*, 36 BRITISH J. CRIMINOLOGY 299–318 (1996).

2. Should Credit Card Issuers Reissue Cards in Response to a Data Breach? Uncertainty and Transparency in Metrics for Data Security Policymaking⁸

2.1. Introduction

In recent years, there has been growing interest in the economic analysis of information security problems—including, in particular, the study of data breaches. Scholars have investigated the impact of data breaches on the stock market valuations of breached firms, the relationship between security investments and the frequency of breaches, and the role of market competition in predicting the probability of a breach. Less explored, however, has been the issue of how private choices by credit card issuers affect the public costs of breach. After a breach of credit cards is disclosed, the financial institutions that issued those cards can either immediately cancel and reissue those cards or instead wait until someone attempts to use the card data for fraud. Reissuing cards can be expensive and potentially wasteful because many cards impacted in a breach may never be used for fraud. But not reissuing cards increases the risk of credit card fraud, which incurs costs to issuers, merchants, and cardholders. No fraud-monitoring program can prevent all fraud. Although issuers may evaluate the internal risks and benefits of reissuing, to my knowledge no published study has attempted to measure the overall societal benefits of each option when costs external to the issuers are considered.

In this chapter, I empirically investigate the social (i.e., aggregate) costs and benefits of reissuing breached cards immediately versus waiting until card fraud is attempted. I analyze “first-order” costs: those costs that are direct results of reissuing cards or leaving them in circulation despite possible compromise. The analysis focuses on societal costs and benefits rather than costs and benefits to issuers.

Although the costs and sources of identity theft are well researched, the connection between identity theft and data breach is not as well understood, nor is quality data available on data breach or its resulting harms. My analysis therefore estimates, based on publicly-available data sources of varying quality, the number of credit cards exposed in data breaches, the cost of identity theft, and the extent to which identity theft is traceable to breaches of credit card data. I analyze public information about reported credit card breaches with known record counts to extrapolate an estimate of unknown records that would also have been exposed. I address uncertainty through parameterization, Monte Carlo analysis, and sensitivity analysis.

This chapter makes two contributions to the literature. First, it confirms that the first-order costs of automatically reissuing cards may be lower than waiting until fraud is attempted.

⁸ This chapter was originally published as James T. Graves, Alessandro Acquisti & Nicolas Christin, *Should Credit Card Issuers Reissue Cards in Response to a Data Breach?: Uncertainty and Transparency in Metrics for Data Security Policymaking*, 18 ACM TRANSACTIONS ON INTERNET TECH. 54 (2018), <https://doi.org/10.1145/3122983>. Citations have been updated where possible, but otherwise, the references and data were current as of early 2016.

Second, it illustrates where improved access to quality data sources is most needed. These results are limited by reliance on publicly available information about data breach and identity theft. Some of this information is excellent, but much of it is not. The extent to which the model is sensitive to different data sources may serve as a guide for where resources could most usefully be spent to improve understanding of the causes of data breach.

Despite these limitations, the result is fairly robust to the tremendous uncertainty in the model. Although the range estimation results in a two order-of-magnitude difference in the estimated cost of fraud if cards are not reissued, the Monte Carlo analysis shows roughly a 91% probability that societal losses would be lower if cards are reissued.

Section 2.2 presents background information placing my research in the context of previous work studying the economics of data breach and cybercrime. Section 2.3 describes the methodology and model. Section 2.4 explains the data I used for the parameters of the model. Section 2.5 presents the analysis of the data. Section 2.6 discusses the implications and some of the limitations of my research.

2.2. Background

Credit card payments rely on relationships between five parties: cardholders, merchants, issuing banks, acquiring banks, and card associations.⁹ Figure 1 illustrates this structure. An acquiring bank (or “acquirer”) is the merchant’s bank; the issuing bank (or “issuer”) is the bank with whom the cardholder has a revolving credit account. The card associations (e.g., MasterCard, Visa, American Express, or Discover) are networks of financial institutions that set rules governing transactions. In the case of American Express and Discover, the card network and issuer are usually the same.

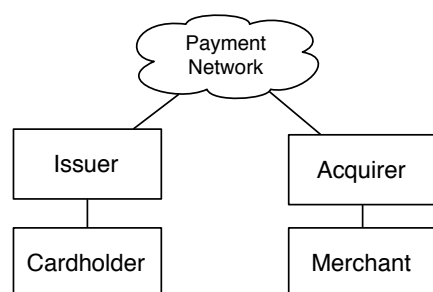


Figure 1: Credit card payment network structure

In simplified form, a credit card transaction works as follows. When a cardholder presents a card for payment at a merchant, the merchant passes the card information and authorization request to its acquiring bank, which forwards the request to the cardholder’s issuing bank. The issuer authorizes or rejects the transaction. If the transaction is authorized, the issuer transfers funds from its payment network account to the acquirer’s payment network

⁹ Adam J. Levitin, *Private Disordering: Payment Card Fraud Liability Rules*, 5 BROOK. J. CORP. FIN. & COM. L. 1, 10–14 (2010).

account. This is called “capture.” Finally, the transaction is “settled” when the acquirer credits the merchant’s account.

In the United States, issuers bear the initial risk of loss from credit card fraud from card-present transactions but the contractual relationships between issuers, the card brands, merchants, and the merchants’ acquiring banks allow those losses to be shifted to merchants that have violated the card brand Operating Regulations by not following prescribed security measures. In most states, however, loss-shifting is available only for fraudulent charges. Issuers bear all the operational costs of reissuing cards and have had little success in lawsuits to recoup these costs from breached merchants. But issuers who sue cannot recover damages they could have avoided. If an issuer could have reduced fraudulent charges to an exposed card by canceling and reissuing that card but did not, the issuer may not be able to recover the cost of those charges if they could have been avoided. Conversely, if the total amount of fraudulent charges that result from a breach are lower than the cost of reissuing the cards, reissuing would be failing to mitigate damages.

In at least one case, merchants have used the fact that an issuer reissued cards and lacked fraud monitoring processes to claim that issuers did not mitigate damages. In the consolidated putative class-action lawsuit resulting from the breach at TJX, one of the retailer’s defenses was that by “unnecessarily and unreasonably automatically canceling and reissuing their customers’ debit cards in response to the data compromise” and by not using fraud monitoring, some of the plaintiffs had either failed to mitigate damages or were contributorily negligent.¹⁰

The cost of reissuing cards is not the only incentive affecting an issuer’s decision whether to reissue. Maintaining cardholder loyalty may be an even more important incentive for issuers to reissue cards even when the cost of doing so might be greater than anticipated fraud. And evidence suggests that issuers do often reissue cards after a breach even if they will not be able to recover the costs of doing so. But the tension between the losses issuers can recover, the operational costs that issuers generally cannot recover, and the obligation to minimize losses raises legal and policy questions. Should the law recognize reissuing costs as reimbursable losses? Is it more societally beneficial to immediately reissue cards or wait? And, more importantly, do we even have the data needed to answer that question or many other public policy questions involving tradeoffs of data security choices?

This chapter tries to answer those questions by building on the literature on the economics of information security, particularly that concerned with data breach. This literature seeks to understand the scope of data breaches, their cost, and the effectiveness of interventions to reduce their impact. Data about breaches has also been used to analyze the economics of security investments more generally.

The full extent of data breaches is difficult to measure. There is currently no comprehensive, openly accessible database of data breaches. The Privacy Rights Clearinghouse (PRC), whose Chronology of Data Breaches is one of the most commonly used data sets for

¹⁰ Answer to Plaintiffs’ Consolidated Class Action Complaint at 22, *In re TJX Retail Security Breach Litigation*, 527 F. Supp. 2d 209 (D. Mass. 2007) (No. 07-10162-WGY).

breach analysis, has warned that its chronology is not a complete list of all breaches.¹¹ The Identity Theft Resource Center publishes annual data breach reports but its list of breaches is not available online.¹² The Open Security Foundation was one of the first to create a database of breaches, but its data, which was once free to download, is no longer available to the general public and its site has since gone dark.¹³ A few states publish lists of the data breaches reported to their attorneys general or other authorities, but these lists include only breaches that affect residents of those states. Three states—Maine, Maryland, and New Hampshire—include estimates of the number of their states’ residents who were affected by each breach (if reported by the organization that was breached). The data from these states might be analyzed in conjunction with the PRC database to obtain a more complete picture of the extent of data breaches.

The distribution of data breaches is heavy-tailed: a few extremely large breaches of millions of records have gotten lots of attention, but most breaches are much smaller. One statistical model predicts, for example, a 31% chance per year of a breach of 10 million records or more in the United States.¹⁴

Early efforts to measure the cost of data breaches were based on surveys. Although they suffer from numerous problems, surveys continue to be popular among industry analysts.¹⁵ One of the first surveys was the Computer Security Institute’s Computer Crime and Security Survey.¹⁶ The most notorious survey of the cost organizations incur after a breach may be the Ponemon Group’s annual study.¹⁷ The Ponemon study has been criticized for methodological issues and a simplistic per-record cost figure that does not accurately reflect costs but invites facile citation by the popular press, product vendors, and security consultants.¹⁸ Verizon’s Data

¹¹ *Chronology of Data Breaches: FAQ*, PRIVACY RIGHTS CLEARINGHOUSE, <https://www.privacyrights.org/chronology-data-breaches-faq> (last visited Nov 6, 2015).

¹² See *Data Breaches*, IDENTITY THEFT RESOURCE CTR., <http://www.idtheftcenter.org/id-theft/data-breaches.html> (last visited Nov 3, 2015).

¹³ <http://datalossdb.org/> (last visited Nov 16, 2015).

¹⁴ Benjamin Edwards, Steven Hofmeyr & Stephanie Forrest, *Hype and Heavy Tails: A Closer Look at Data Breaches*, in 14TH ANN. WORKSHOP ON ECON. INFO. SECURITY (2015), https://econinfosec.org/archive/weis2015/papers/WEIS_2015_edwards.pdf.

¹⁵ Jay Heiser, *Can Information Security Surveys Be Trusted?*, TECHTARGET.COM (2002), <http://searchsecurity.techtarget.com/feature/Can-information-security-surveys-be-trusted> (last visited May 27, 2016); Julie J. C. H. Ryan & Theresa I. Jefferson, *The Use, Misuse, and Abuse of Statistics in Information Security Research*, in PROC. OF 24TH ANN. NAT’L ASEM (2003); ADAM SHOSTACK & ANDREW STEWART, THE NEW SCHOOL OF INFORMATION SECURITY 46 (2008).

¹⁶ Computer Security Institute, *1997 CSI/FBI Computer Crime and Security Survey*, 3 COMPUTER SECURITY ISSUES AND TRENDS (1997).

¹⁷ PONEMON INSTITUTE, *2015 Cost of Data Breach Study: Global Analysis* (2015), <http://www-03.ibm.com/security/data-breach/>.

¹⁸ Robert Hackett, *The Hotly Disputed Black Magic of Data Breach Cost Estimates*, FORTUNE (Apr. 24, 2015), <http://fortune.com/2015/04/24/data-breach-cost-estimate-dispute/>; Jay Jacobs, *Analyzing Ponemon Cost of Data Breach*, DATA DRIVEN SECURITY (Dec. 11, 2014), <http://datadrivensecurity.info/blog/posts/2014/Dec/ponemon/>; Adam Shostak, *A Critique of Ponemon Institute Methodology for “Churn”*, NEW SCH. OF INFO. SECURITY (Jan. 25, 2011), <http://newschoolsecurity.com/2011/01/a-critique-of-ponemon-institute-methodology-for-churn/>.

Breach Investigations Report (DBIR),¹⁹ which added an estimate of the cost of breaches for the first time in its 2015 edition, argues that the cost of a breach is best modeled by a nonlinear function of the number of records breached.

A popular empirical method of estimating the cost of breaches to firms is to measure the effect of a breach announcement on stock prices. One of the earliest studies to use this approach found an average abnormal drop in stock price of 4.5% over three days in the 22 security breach events in the authors' sample.²⁰ Other studies have found similar short-term post-breach drops in market value²¹ and profits.²² Although recent research still finds a statistically significant short-term drop in stock prices, the effect has gone down over time, perhaps because breaches have become more commonplace.²³ In contrast to the short-term hit on stock price, most firms do not appear to suffer long-term drops in market value after a breach.²⁴ And the effect of different types of breach is not uniform. In a study of 43 security breaches from 1995–2000, breaches related to confidential information were associated with drops in stock prices but breaches that “largely affected the information infrastructure itself” were not.²⁵

Another area of data breach economics research focuses on the effects of data breach notification laws. Lenard and Rubin have argued that the costs of these laws outweigh their benefits.²⁶ Extrapolating from limited public data on the cost and incidence of identity theft, they concluded that the expected benefit from notifying consumers of a data breach was in the range of \$7.50 to \$10—lower than the costs they listed from notification, which included \$10–\$20 per card to reissue cards and \$2 per card to send notification letters. But even if notification

¹⁹ VERIZON ENTERPRISE SOLUTIONS, 2015 DATA BREACH INVESTIGATIONS REPORT (2015), <http://www.verizonenterprise.com/DBIR/2015/>.

²⁰ Ashish Garg, Jeffrey Curtis & Hilary Halper, *Quantifying the Financial Impact of IT Security Breaches*, 11 INFO. MGMT. & COMPUTER SECURITY 74 (2003).

²¹ Alessandro Acquisti, Allan Friedman & Rahul Telang, *Is There a Cost to Privacy Breaches? An Event Study*, in PROC. OF 27TH INT'L CONF. ON INFO. SYS. (2006); Katherine Campbell et al., *The Economic Cost of Publicly Announced Information Security Breaches: Empirical Evidence from the Stock Market*, 11 J. COMPUTER SECURITY 431 (2003); Huseyin Cavusoglu, Birendra Mishra & Srinivasan Raghunathan, *The Effect of Internet Security Breach Announcements on Market Value: Capital Market Reactions for Breached Firms and Internet Security Developers*, 9 INT'L J. ELECTRONIC COM. 70 (2004); Kevin M. Gatzlaff & Kathleen A. McCullough, *The Effect of Data Breaches on Shareholder Wealth*, 13 RISK MGMT. & INS. REV. 61 (2010); Sanjay Goel & Hany A. Shawky, *Estimating the Market Impact of Security Breach Announcements on Firm Values*, 46 INFO. & MGMT 404 (2009); Lawrence A. Gordon, Martin P. Loeb & Lei Zhou, *The Impact of Information Security Breaches: Has There Been a Downward Shift in Costs?*, 19 J. COMPUTER SECURITY 33 (2011).

²² Kholekile L. Gwebu, Jing Wang & Wenjuan Xie, *Understanding the Cost Associated with Data Security Breaches*, in PROC. 19TH PACIFIC ASIA CONF. ON INFO. SYS. (2014); Kweku-Muata Osei-Bryson, Myung Ko & Humayun Zafar, *Financial Impact of Information Security Breaches on Breached Firms and Their Non-Breached Competitors*, 25 INF. RESOURCE MGMT. J. 21 (2012).

²³ Gordon, Loeb & Zhou, *supra* note 21.

²⁴ Karthik Kannan, Jackie Rees & Sanjay Sridhar, *Market Reactions to Information Security Breach Announcements: An Empirical Analysis*, 12 INT'L J. ELEC. COM. 69 (2007).

²⁵ Campbell et al., *supra* note 12.

²⁶ THOMAS M. LENARD & PAUL H. RUBIN, AN ECONOMIC ANALYSIS OF NOTIFICATION REQUIREMENTS FOR DATA SECURITY BREACHES, EMORY L & ECON. RES. PAPER 05-12 (2005), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=765845.

laws increase costs to firms, they may reduce overall social costs by causing firms and consumers to improve their levels of data security care.²⁷ Romanosky et al., for example, found that notification laws may reduce identity theft by about 6%.²⁸

Statistics about data breaches have also been used as inputs to empirical analyses of the effectiveness of data security investments. Miller and Tucker, for example, found no evidence that adoption of encryption software among hospitals reduced the number data breaches.²⁹ To the contrary, they found that public announcements of certain types of data breach actually increased. Gaynor et al. used an analysis of breach data to reach the surprising conclusion that hospitals in competitive healthcare markets seem to be worse at protecting patient data than those in non-competitive markets.³⁰ Kwon and Johnson applied a proportional hazard model to breach disclosures by 281 healthcare organizations to find that security measures appear to be more effective when adopted voluntarily instead of being forced by regulation.³¹

Before the publication of the article on which this chapter is based, there was little to no academic literature on how financial institutions decide whether to reissue cards after a breach, a decision process that the institutions treat as proprietary. The sole source I could find, other than news reports, was a 2008 study by the state of Maine surveying banks' responses to two major data breaches in that state.³² That study reported that issuers reissued 78% of cards during the period covered by the survey.

2.3. Methodology

This chapter is an attempt to estimate and compare the aggregate first-order net social (i.e., aggregate) costs that result from decisions by issuers who, upon a credit card breach, face a choice between reissuing cards or waiting. "Social costs" include the total costs regardless of who incurs them (although any benefit gained by criminals is ignored). The term "first-order costs," as used here, refers to those costs that are direct results of reissuing cards or leaving them in circulation despite possible compromise. These can include the costs (including overhead) of mailing replacement cards, time spent by merchants and consumers responding to having cards reissued, or, for cards that are not reissued, the expected cost of fraud on those cards. First-order

²⁷ Sasha Romanosky, Alessandro Acquisti & Richard Sharp, *Data Breaches and Identity Theft: When is Mandatory Disclosure Optimal?* in TECH. POL'Y RES. CONF. 2010 (2010), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1989594.

²⁸ Sasha Romanosky, Rahul Telang, & Alessandro Acquisti, *Do Data Breach Disclosure Laws Reduce Identity Theft?*, 30 J. POL'Y ANAL. & MGMT. 2 (2011).

²⁹ Amalia R. Miller & Catherine Tucker, *Encryption and Data Loss*, in 2010 WORKSHOP ON ECON. INFO. SECURITY (2010), http://weis2010.econinfosec.org/papers/session1/weis2010_tucker.pdf

³⁰ Martin S. Gaynor, Muhammad Zia Hydari, and Rahul Telang, *Is Patient Data Better Protected in Competitive Healthcare Markets?* in 2012 WORKSHOP ON ECON. INFO. SECURITY (2012), https://econinfosec.org/archive/weis2012/papers/Gaynor_WEIS2012.pdf.

³¹ Juhee Kwon & M. Eric Johnson, *An Organizational Learning Perspective on Proactive vs. Reactive Investment in Information Security*, in 2011 WORKSHOP ON ECON. INFO. SECURITY (2011).

³² ME. BUREAU OF FIN. INSTS., MAINE DATA BREACH STUDY (2008), <http://www.state.me.us/pfr/financialinstitutions/reports/index.htm>.

costs are distinguished from “second-order” effects, which are indirect costs that occur over time or that are in some other sense a step removed from the immediate costs.³³ Because I am interested in aggregate social costs, who incurs the cost of fraud is less critical to the model than is the total amount of that fraud.

I restrict the scope of my analysis in a number of ways. First, I concentrate on credit cards rather than debit cards or other payment instruments that have different authentication structures and risk profiles from credit cards. Second, the analysis is specific to the United States. Third, I concentrate on overall social costs, largely because, to my knowledge, no publicly available data exists that would enable an analysis of the allocation of those costs between parties. Fourth, I focus specifically on existing-account credit card fraud as the primary cost of credit card fraud. This is a subtype of identity theft in which victims’ existing credit cards are used for unauthorized charges. Credit card data is unlikely to facilitate other forms of identity theft such as new-account fraud (in which new accounts are opened using the victim’s identity) because opening a new account requires more than a credit card.

I use the following model of first-order costs:

$$\sum_k r c_{i_k} + (1 - r) \rho_k f_k \quad (1)$$

The model sums, over each affected card k , the costs related to that card, with the following terms:

- r : Binary variable where $r = 1$ if the card is reissued and $r = 0$ if not
- c_{i_k} : Cost of reissue for issuer i_k of card k
- ρ_k : Probability that card k will be used fraudulently
- f_k : Amount of fraud if the card is used fraudulently

The model omits potential costs to cardholders and merchants of issuers reissuing cards because I assume that these costs are relatively small. A canceled and reissued credit card used for recurring payments may lead to a merchant having to contact customers to obtain new payment information, but these processes are generally automated and inexpensive.³⁴ The costs to cardholders come from the value of time spent responding to the cancelation—for example, updating auto-pay accounts to use the new card number. Issuers can minimize these costs by sending replacement cards before canceling outstanding cards, but cardholders do sometimes miss or ignore the payment cards or are traveling when the replacements are made.³⁵

³³ See *infra* Section 2.6.1.

³⁴ *Pricing*, AUTHORIZE.NET (2016), <http://www.authorize.net/solutions/merchantsolutions/pricing/>; *Insights: Authorization Fee*, CAYAN (2010), <https://cayan.com/glossary/authorization-fee>.

³⁵ Eric Stark, *Computer Hackers Are Stealing Bank Card Information, but There is Protection and Some Banks Have Been Aggressive*, SUNDAY NEWS (Lancaster, Pa.), at 1 (July 11, 2004).

The model also assumes that fraud losses are zero if cards are reissued. Although it may be possible to use canceled cards fraudulently if a merchant is not vigilant about clearing authorization before goods or services have been rendered, I assume that the overall loss from these pre-authorization transaction losses is negligible.

Because no data is publicly available for ρ_k , it must be estimated. I use the following equation:

$$\rho_k = (1 - \delta) \left[1 - \left(1 - \frac{vb}{\theta(n_d\gamma(1 - \lambda) + n_u)} \right)^{1/a} \right] \quad (2)$$

This part of the model relies on the following parameters:

- n_d : Number of payment card records affected each year in disclosed data breaches for which the number of records affected is made public
- n_u : Number of payment card records affected each year in disclosed data breaches for which the number of records affected is either unknown or not made public
- θ : Scaling factor to account for payment card records exposed in breaches that are either undiscovered, undisclosed, or not included in the data to which I have access
- λ : Breached credit cards that are immediately reissued by issuing institutions, as a proportion of all breached credit cards
- γ : Credit cards as a proportion of all payment cards. This parameter captures the fact that breach disclosures may use “credit cards” to refer to payment cards generally, whereas the model focuses solely on credit cards
- v : Number of people victimized by existing-account credit card fraud per year
- b : Proportion of existing-account credit card fraud attributable to data breach as the method by which the card data was obtained
- a : Average number of credit cards per cardholder. This parameter allows us to use per-person data on the cost of credit card fraud in the model of the cost per card
- δ : Reduction in the probability of fraud achieved by an issuer flagging breached cards in its fraud detection algorithms

Equation (2) estimates the probability of card misuse following a breach as a function of the number of existing-account credit card fraud incidents attributable to data breach (vb); the total number of credit card records exposed in breaches each year, including those that are not included in breach databases either because the scope of a breach was unknown or because the breach was not discovered or publicly disclosed ($\theta(n_d\gamma(1 - \lambda) + n_u)$); the effectiveness of fraud detection algorithms (δ); and the number of credit cards per person (a).

I estimate the amount of fraud if a card is misused as:

$$f_k = c_{m_k} + t_k c_{t_k} + c_{i_k} \quad (3)$$

The parameters in this part of the model are:

c_{m_k} : Monetary cost of existing account credit card fraud per incident

t_k : Time (in hours) spent responding to existing account credit card fraud by cardholders

c_{t_k} : Cost of cardholder time (per hour)

I include c_{i_k} , first used in expression (1), to capture the cost of canceling and reissuing cards that have been used for fraud.

Substituting the formulas for ρ_k and f_k in expression (1) results in the following model that includes all parameters:

$$\sum_k r c_{i_k} + (1 - r)(1 - \delta) \left[1 - \left(1 - \frac{vb}{\theta(n_d \gamma(1 - \lambda) + n_u)} \right)^{1/a} \right] c_{m_k} + t_k c_{t_k} + c_{i_k} \quad (4)$$

This calculation assumes that the card records exposed in breaches are unique—i.e., that two different breach events do not expose the same credit card record. Overlap between breaches would reduce the total number of credit card records exposed. This assumption seems reasonable given the current common (but not universal) practice of reissuing credit cards potentially exposed in a breach. I also assume that the same breached card is not victimized twice (where a “victimization” may include multiple fraudulent charges). This follows from the assumption that fraudulently used cards will immediately be cancelled and reissued once that fraud is detected.

The calculations also use annual averages even though the number of cards exposed in data breaches varies widely from year to year. Using annual averages reflects the assumption that both collection and misuse of credit cards occurs over time. Although a massive breach may be announced on a certain date, access to the data may have occurred over weeks or months. Thus, it seems to make sense to smooth this data by considering annual averages and not focusing on individual yearly totals.

2.4. Data

This section discusses the data sources for each of the parameters presented above and describes the ranges and point estimates used for each parameter.

2.4.1. The Cost of Reissuing Cards

Three types of data sources shed light on the cost of reissuing cards: news reports, lawsuits, and a state government survey. News reports have quoted figures from issuers and other industry sources; these estimates range from \$3 to \$25 per card.³⁶ Lawsuits filed by issuers seeking to recover the cost of reissuing cards claim losses from reissuing of \$5 to \$20 per card, with some evidence that economies of scale reduce the per-card cost when an institution must reissue more cards.³⁷ The state of Maine conducted a survey that found a cost of \$4.72 per card reported by issuers in that state.³⁸

Considering these sources as a whole, it appears that the cost (c_{i_k}) is between \$5–\$25 per card. Because the cost for most issuers seems to be \$10 or less, I use \$10 as a point estimate.

2.4.2. The Probability of Credit Card Misuse Following a Breach

To the best of my knowledge, no publicly available data exists on the probability that a credit card affected in a data breach will be used for fraud. I estimate that probability (ρ_k) by multiplying the number of annual incidents of existing-account credit card fraud (v) by the proportion of those incidents in which the credit card data was obtained using data breach (b) then dividing that by the total number of credit cards exposed in data breaches each year ($\theta(n_d\gamma(1 - \lambda) + n_u)$).

I assume that the majority of issuers already use some form of fraud monitoring. This has two implications. First, the marginal cost to monitor a card that has been exposed in a data breach is essentially zero. Setting a flag in an issuer's fraud monitoring system has negligible marginal cost if the database is set up to accommodate such a flag. Second, this assumption implies that the current level of existing-account credit card fraud already reflects the use of fraud monitoring and prevention systems. Flagging a card might improve the probability that attempted fraud will be detected and prevented—at some risk of additional false positives—but the baseline probability of fraud does not rely on the effectiveness of current fraud-monitoring processes.

³⁶ America's Community Bankers, *ACB Data Breach Survey Highlights Need for Action by Card Networks and Congress*, PR NEWswire (Feb. 7, 2007); Maria Aspan & Clare Baldwin, *Sony Breach Could Cost Card Lenders \$300 Mln*, REUTERS (Apr. 29, 2011); Chris Churchill, *TJX Reacts to Bank Lawsuit*, TIMES UNION (Albany, N.Y.) (Aug. 30, 2008); Tamara E. Holmes, *Credit Card Fraud and ID Theft Statistics*, CREDITCARDS.COM (Sept. 16, 2015), <http://www.creditcards.com/credit-card-news/credit-card-security-id-theft-fraud-statistics-1276.php>; SHIRLEY W. INSCOE, AITE GROUP, GLOBAL CONSUMERS REACT TO RISING FRAUD: BEWARE BACK OF WALLET (2012); Mark Jewell, *IDs Are a Steal; Thieves Looking for Credit Numbers Set Their Sights on Big Targets*, COLUMBIAN (Vancouver, Wash.), at E (Aug. 23, 2004); Andrew Johnson, *Card Fraud Risk Low from Breach at Citi*, AM. BANKER 10 (June 10, 2011); Ann Ravana, *Banks Start Credit Card Reissue*, BANGOR DAILY NEWS 4 (Feb. 8, 2007); Stark, *supra* note 35.

³⁷ See, e.g., *Pa. State Emps. Credit Union v. Fifth Third Bank*, 398 F. Supp. 2d 317, 322 (2005) (stating that PSECU canceled 20,029 cards at a total cost of \$98,128.13).

³⁸ ME. BUREAU OF FIN. INST., *supra* note 32.

2.4.2.1. Payment Cards Exposed in Data Breaches with Record Counts

The number of records exposed in data breaches is uncertain for three reasons. First, only breaches that are discovered can be counted. Second, not all discovered breaches are publicly disclosed. And third, even when a breach has been detected and reported, it may not be possible to determine how many records were exposed. The model contains parameters for three types of breached payment card records: those that are publicly disclosed with estimated record counts (n_d), those that are disclosed with unknown record counts (n_u), and a scaling factor to account for undetected breaches (θ).

The record count I use in the model is based on a detailed analysis of the Privacy Rights Clearinghouse (PRC) database.³⁹ I calculated the number of payment cards potentially exposed by downloading the PRC database, filtering based on the use of the word “card” in the description field, and manually categorizing each entry, based on its description, as having potentially exposed full unencrypted payment card numbers or not. Thus, I did not include breach events that were described as having exposed only partial or encrypted payment card numbers. I did not, however, filter out breach events in which card numbers were exposed without other “full track” data such as expirations dates. Recent work by Ali et al.⁴⁰ shows that due to different online merchants using different fields for verifying card transactions, it is easy for an attacker with just a card number to discern all the other information needed to use the number for fraud. I also omitted events disclosed in 2005 because breach reporting was still new and the 16 events reported for that year were probably non-representative. Where necessary, I updated PRC’s record counts to reflect only the number of payment cards believed to have been exposed. My analysis of the PRC database yielded a list of 579 breach events from 2006 through the end of 2014. Of those events, 269 included record counts (46%).

I supplemented this data with information from the Maine, Maryland, and New Hampshire data breach lists.⁴¹ I used these states’ lists to add breaches that were not included in the PRC database and to estimate record counts for breaches where PRC did not have those numbers. This added 179 breach events to the database—including 34 with overall record counts—for a total of 758, of which 303 included record counts (40%). Those records total 378 million payment card accounts over nine years. I use that average of about 42 million cards per year as the point estimate. Because the number of reported records is relatively well known, I use the narrow range of 39 million to 45 million cards per year for this parameter (n_d).

³⁹ *Data Breaches*, PRIVACY RIGHTS CLEARINGHOUSE, <https://www.privacyrights.org/data-breaches> (last visited Nov 3, 2015).

⁴⁰ *Does the Online Card Payment Landscape Unwittingly Facilitate Fraud?*, 15 IEEE SECURITY & PRIVACY 78 (2017).

⁴¹ *Privacy, Identity Theft and Data Security Breaches*, ME. ATT’Y GEN., http://www.state.me.us/ag/consumer/identity_theft/index.shtml (last visited Nov 16, 2015); *Maryland Information Security Breach Notices*, MD. ATT’Y GEN., <http://www.marylandattorneygeneral.gov/Pages/IdentityTheft/breachnotices.aspx> (last visited Nov 16, 2015); *Security Breach Notifications*, N.H. ATT’Y GEN., <http://doj.nh.gov/consumer/security-breaches/> (last visited Nov 16, 2015).

2.4.2.2. Payment Card Records Exposed in Data Breaches without Record Counts

I used two methods to estimate the total number of records exposed in breach events without record counts (n_u). First I used linear regressions to predict overall record counts from the number of residents of Maine, Maryland, and New Hampshire that were affected. This gave estimates for an additional 231 breach events at a total of about 630,000 records per year.

For the remaining events in the database, I extrapolated using a weighted estimate based on the typical number of records exposed for each type of data breach. I excluded the TJX, Heartland, Target, and Home Depot breaches because I believe it unlikely that any of the disclosed breaches with unknown record counts could have exposed records on the order of the tens of millions or hundreds of millions of records exposed in those four breaches. I also excluded one insider breach at Fidelity because it appears to be an extreme outlier: the 8.5 million records compromised in that breach were two orders of magnitude larger than any other insider breaches with known record counts and three orders of magnitude larger than the average in that category when the Fidelity breach is excluded.

Based on the weighted average, I estimate that the 224 breaches with unknown record counts from 2006 through 2014 have exposed about 2.4 million accounts per year. This estimate may still be too high because it includes nine other breaches in which at least a million records were believed to have been affected. Excluding these breaches gives a weighted estimate of about 580,000 records per year from unreported breaches. I use a point estimate of $n_u = 2.1$ million unknown breached records per year, which is derived from the linear regression estimate of 630,000 added to the midpoint between the 580,000 and 2.4 million estimates from the weighted average extrapolation. But the range is wide—1 million to 10 million cards per year—because of the uncertainty surrounding the number of records in breaches for which record counts were not disclosed.

2.4.2.3. Payment Card Records Exposed in Undetected or Undisclosed Breaches

It is impossible to know how many breaches are not detected. There is, however, plenty of speculation. For example, one security product vendor (with the possible biases that implies) claims that 85% of data breach events are undetected.⁴² The number of data breaches that went undetected for months or years suggests that there have probably been other breaches that were not detected at all.⁴³ Perhaps more enlightening are the controlled penetration tests conducted at

⁴² Gaby Friedlander, *Why 85% of Data Breaches Are Undetected*, OBSERVEIT (July 16, 2014), <http://www.observeit.com/blog/why-85-percent-data-breaches-undetected>.

⁴³ Steve Gold, *Home Depot Card Data Breach Undetected for Four Months*, SC MAGAZINE UK (Sept. 22, 2014), <http://www.scmagazineuk.com/news/home-depot-card-data-breach-undetected-for-four-months/article/372794/>; Sean Micheal Kerner, *UPS Discloses Data Breach that Went Undetected for Months*, EWEEK (Aug. 21, 2014), <http://www.eweek.com/blogs/security-watch/ups-discloses-data-breach-that-went-undetected-for-months.html>; Nathaniel Popper, *Breach at Neiman Marcus Went Undetected from July to December*, N.Y. TIMES, at B1 (Jan. 17, 2014); *OPM Data Breach Undetected for a Year*, PYMNTS.COM (June 22, 2015), <http://www.pymnts.com/news/2015/opm-data-breach-undetected-for-a-year/>.

government agencies. In Fiscal Year 2011, 49% of those intrusions were detected.⁴⁴ That increased to 73% in Fiscal Year 2013.⁴⁵ Although these numbers are the results of controlled tests against specific goals and agencies, they offer a general idea of the extent to which intrusions are detected overall.

The other type of unknown included in θ is the number of records in breaches that are detected but not disclosed. Some surveys attempt to measure a similar variable. For instance, one survey of “malware analysts” found that 57% claimed that their organizations had not disclosed data breaches.⁴⁶ There are serious methodological problems with this figure, such as the difficulty in translating the number of analysts to a number of records and lack of clarity as to the definition of a “breach,” but better information does not seem to be available.

Because this number is subject to much uncertainty, I use the broadest range that seems plausible. I assume that undetected and undisclosed breaches expose between one-fourth and three times as many records as are exposed in detected breaches, with a conservative point estimate of $\theta = 1.75$. This potentially overestimates the number of records that are exposed, which could be the case if, for example, undetected or undisclosed breaches tend to be smaller than those that are detected and disclosed.

2.4.2.4. Proportion of Breached Cards that are Immediately Reissued

The next factor needed to calculate ρ_k is the percentage of breached credit cards that are reissued before fraud occurs (λ). A 2008 Maine study of banks’ responses to data breach incidents found that issuers reissued 78% of cards during the period covered by the survey, including 84% of accounts affected in the TJX breach and 77% of those affected in the Hannaford breach.⁴⁷ Another source claims that “nearly 90 percent of card breach victims in 2014 received replacement credit cards.”⁴⁸ I therefore assume that issuers re-issue between roughly 80% and 95% of cards, with a point estimate of 87.5%.

2.4.2.5. Credit Cards as a Proportion of Payment Cards

According to data from the Statistical Abstract of the United States and the Nilson Report, credit cards (excluding store cards, oil company cards, and other non-general-purpose cards) have decreased as a percentage of all payment cards from 58% in 2008 to 47% in 2014.⁴⁹

⁴⁴ OFFICE OF MGMT. & BUDGET, FISCAL YEAR 2012 REPORT TO CONGRESS ON THE IMPLEMENTATION OF THE FEDERAL INFORMATION SECURITY MANAGEMENT ACT OF 2002 (2013).

⁴⁵ OFFICE OF MGMT. & BUDGET, ANNUAL REPORT TO CONGRESS: FEDERAL INFORMATION SECURITY MANAGEMENT ACT (2014).

⁴⁶ ThreatTrack Security, *Malware Analysts Have the Tools They Need, but Challenges Remain*, BANKINFOSECURITY.COM (June 5, 2014), <http://www.bankinfosecurity.com/whitepapers/malware-analysts-have-tools-they-need-but-challenges-remain-w-1026>.

⁴⁷ ME. BUREAU OF FIN. INST., *supra* note 32.

⁴⁸ Holmes, *supra* note 36.

⁴⁹ HSN CONSULTANTS, INC., THE NILSON REPORT: GENERAL PURPOSE CARDS—U.S. 2012 (2013); HSN CONSULTANTS, INC., THE NILSON REPORT: GENERAL PURPOSE CARDS—U.S. 2013 (2014); HSN CONSULTANTS,

Under the assumption that the proportion of credit cards to payment cards in breaches is the same as the proportion in general circulation, I use these values as the range of the model for γ , with a point estimate of 52%.

Table 1 summarizes the ranges and point values for n_d , n_u , θ , λ , and γ . Taking the high and lows of this range, I estimate that between about 2.4 million and 60.9 million non-reissued credit card records are exposed in data breaches annually, with a point estimate of 8.5 million cards.

Table 1: Estimated total number of credit card records exposed in data breach per year

<i>Description</i>	<i>Low</i>	<i>Point</i>	<i>High</i>
Payment card records reported lost in data breaches per year (n_d) (mil)	39	42	45
Est. records per year in breaches with unknown record counts (n_u) (mil)	1.00	2.10	10.00
Scaling factor to account for unreported or undetected breaches (θ)	1.25	1.75	4.00
Portion of breached cards reissued (λ)	0.95	0.88	0.80
Credit cards as a proportion of breached payment cards (γ)	0.47	0.52	0.58
Total credit card records exposed in all breaches per year (mil)	2.40	8.50	60.90

2.4.2.6. Number of People Affected by Existing-Account Credit Card Fraud

The Department of Justice’s Bureau of Justice Statistics (BJS) has included identity theft questions in its annual National Crime Victimization Survey (NCVS) since 2004.⁵⁰ These statistics are split out by the nature of the crime; “existing account credit card identity theft” refers to situations in which existing credit cards were used without the cardholder’s authorization.

In 2008, BJS began adding an Identity Theft Supplement (ITS) to the NCVS. The questions in this supplement collected data on identity theft experienced by individuals instead of households. The 2008 supplement asked if respondents had experienced identity theft in the two years prior to the interview. In the 2012 and 2014 surveys, the ITS asked about individual-level identity theft over the previous 12 months. As a result, it is not possible to compare results across the 2005–2010 surveys, the 2008 survey, or the 2012–2014 surveys.⁵¹

Using the 2012 and 2014 per-person data and taking the overall minimum and maximum of the 95% confidence intervals for each year results in a range of 6.8 million to 9.0 million people affected by existing-account credit card fraud each year. I take the average of the 2012 and 2014 point estimates to set $\nu = 8.15$ million people.

INC., THE NILSON REPORT: GENERAL PURPOSE CARDS—U.S. 2014 (2015); U.S. CENSUS, 2011 STATISTICAL ABSTRACT OF THE UNITED STATES, t. 1186, 1187; U.S. CENSUS, 2012 STATISTICAL ABSTRACT OF THE UNITED STATES, t. 1187, 1188.

⁵⁰ ERIKA HARRELL, U.S. DEPT. OF JUSTICE, VICTIMS OF IDENTITY THEFT, 2014 (2015), <http://www.bjs.gov/content/pub/pdf/vit14.pdf>.

⁵¹ ERIKA HARRELL & LYNN LANGTON, U.S. DEPT. OF JUSTICE, VICTIMS OF IDENTITY THEFT, 2012 (2013), <http://www.bjs.gov/index.cfm?ty=pbdetail&iid=5408>.

2.4.2.7. Proportion of Existing Credit Card Fraud Attributable to Breach

Not all credit card fraud is the result of breach. Victims of existing-account credit card fraud who know how their card information was obtained most often say that it was through a stolen wallet or a purse or from someone they know. Breach seems to be a relatively infrequent cause of credit card fraud, but it is uncertain how infrequent. Surveys by Javelin Research, the Identity Theft Research Center, the FTC, and the DOJ's Bureau of Justice Statistics have asked victims of identity theft if they knew how their information was obtained.⁵² Utica College's Center for Identity Management and Information Protection (CIMIP) analyzed the same question (among many others) using federal criminal case data.⁵³ Figure 2 shows the results of these studies.

Most survey respondents did not know how their data was obtained. The responses of those who said that they knew how their data was obtained can legitimately be generalized only if the point of compromise and the victim's knowledge of that point of compromise are uncorrelated. But this may not be true. Some points of compromise are more likely to be known than others. Lost wallets, purses, or thefts alert a cardholder that their cards may have been stolen. Other points of compromise, such as skimmers (devices that surreptitiously record card data at an ATM or point of payment) are unlikely to be recognized. People whose cards are compromised through phishing or spyware will not always know that their cards were obtained in that matter. A data breach, of which a cardholder must be notified in forty-six of fifty states, may be more or less likely to be a known point of compromise.

The ITRC survey is an outlier in this set, with the highest percentage of known points of compromise and the highest percentage of people responding that their data was obtained in a breach. As the ITRC⁵⁴ acknowledges, "[t]his may be due to the fact that ITRC is listed as a victim resource by many entities which have suffered a breach."

⁵² BUREAU OF JUSTICE STATISTICS, NATIONAL CRIME VICTIMIZATION SURVEY: IDENTITY THEFT SUPPLEMENT, 2012 (2014), <http://doi.org/10.3886/ICPSR34735.v1>; HARRELL AND LANGTON, *supra* note 51; IDENTITY THEFT RES. CTR., IDENTITY THEFT: THE AFTERMATH 2009 (2010), <http://www.idtheftcenter.org/ITRC-Surveys-Studies/aftermathstudies.html>; JAVELIN STRATEGY & RESEARCH, 2009 IDENTITY FRAUD SURVEY REPORT: CONSUMER VERSION (2009), https://www.javelinstrategy.com/uploads/files/901.R_Identity_Fraud_Survey_Consumer_Report.pdf; LYNN LANGTON & MICHAEL PLANTY, U.S. DEPT. OF JUSTICE, VICTIMS OF IDENTITY THEFT, 2008 (2010); SYNOVATE, FEDERAL TRADE COMMISSION—2006 IDENTITY THEFT SURVEY REPORT (2007), <https://www.ftc.gov/reports/federal-trade-commission-2006-identity-theft-survey-report-prepared-commission-synovate>.

⁵³ GARY GORDON ET AL., IDENTITY FRAUD TRENDS AND PATTERNS: BUILDING A DATA-BASED FOUNDATION FOR PROACTIVE ENFORCEMENT (2007), <http://www.utica.edu/academic/institutes/cimip/publications/index.cfm>; DONALD J. REBOVICH, KRISTY ALLEN & JARED PLATT, THE NEW FACE OF IDENTITY THEFT: AN ANALYSIS OF FEDERAL CASE DATA FOR THE YEARS 2008 THROUGH 2013 (2015), https://www.utica.edu/academic/institutes/cimip/New_Face_of_Identity_Theft.pdf (last visited Nov 18, 2015).

⁵⁴ IDENTITY THEFT RES. CTR., *supra* note 52.

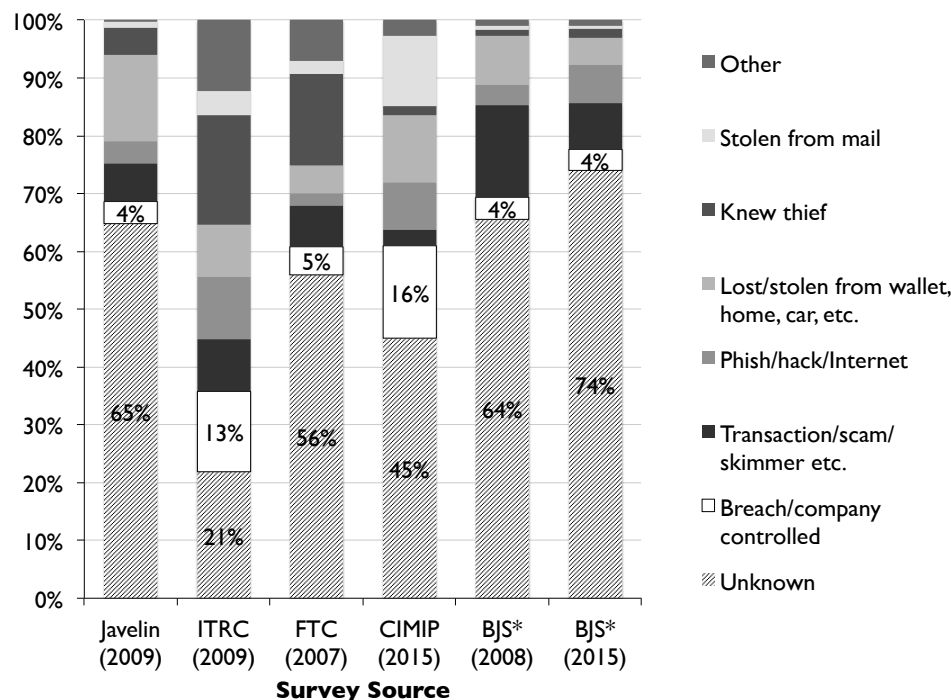


Figure 2: Survey and study results for the number of identity theft victims knew how their data was obtained, and, if so, the point of compromise

Only the BJS survey reported responses for points of compromise specifically for existing-account credit card fraud. None of the other surveys distinguished between forms of identity theft in their reporting. Based on the factors listed above, I use a range of 5% to 15% as the proportion of existing-account credit card fraud in which the card information was obtained in a data breach (b), with a point estimate of 11%. I choose this range to capture, at the low end, either the lowest estimate for breach as a percentage of known points of compromise or the midrange of estimates for breach as a percentage of all compromise, including unknown sources. The high end of the range is just below the ITRC's number, which has a high number of people who believe they know how their information was obtained and the aforementioned potential bias toward identifying breach as the way credit card information was obtained.

2.4.2.8. Reduction in Fraud from Flagging Breached Cards

I assume that flagging exposed cards reduces fraud rates by up to 20%. As discussed at the start of this section, current levels of fraud monitoring are already reflected in existing credit card fraud statistics. Thus, marking a card as potentially exposed can at best improve the effectiveness of fraud monitoring systems somewhat. Unfortunately, information on the effectiveness of fraud monitoring software is treated as proprietary by both issuers and the software vendors. The 0% to 20% range (with a 10% point estimate) for δ therefore represents a best guess.

2.4.2.9. Number of Credit Cards per Cardholder

Converting from the per-person data reported by the BJS to per-card numbers requires an estimate of the number of credit cards per cardholder (a). According to Gallup polls, credit card owners hold an average of about 3.6 to 3.7 credit cards each from 2006–2014.⁵⁵ Surveys conducted by the Federal Reserve Bank of Boston found that cardholders had between 3.8 and 4.0 cards each from 2010–2012.⁵⁶ These numbers include Mastercard, Visa, American Express, and Discover cards but exclude store cards (which are valid only at the stores that issue them), gas company cards, and other specialty cards such as phone cards. I use the high and low end these numbers as an estimated range of 3.6 to 4.0, with a point value in the middle at 3.8.

2.4.2.10. Calculation of ρ_k

Using the parameter values discussed above (which are summarized in Table 2) results in an estimated range for ρ_k of 0.0011 to 0.21, with a point estimate of 0.026.

Table 2: Calculation of the probability of existing-account credit card fraud to an account affected by a breach

<i>Description</i>	<i>Low</i>	<i>Point</i>	<i>High</i>
Number of credit cards exposed (from Table I) (mil)	2.40	8.50	60.90
Number of persons victimized (v) (mil)	6.80	8.15	9.00
Percent of existing-account credit card fraud from breach (b)	5%	11%	15%
Fraud reduction from flagging exposed cards (δ)	0%	10%	20%
Average number of credit cards per cardholder (a)	3.6	3.8	4.0
P(existing-account credit card fraud breach) (ρ_k)	0.0011	0.026	0.21

2.4.3. The Cost of Credit Card Fraud

The cost of an existing-account credit card fraud incident (f_k) has two components: financial losses, including both the loss of value obtained through the fraud and indirect financial costs from responding to the fraud, and the cost of time spent dealing with the fraud.

In most cases, a cardholder should suffer little or no direct out-of-pocket loss from existing-account credit card fraud. Federal law limits cardholder liability to \$50 for unauthorized credit card charges if a lost or stolen card is reported as soon as the loss or theft is discovered.⁵⁷ Visa and Mastercard have voluntary zero-liability policies that further reduce consumer liability

⁵⁵ Art Swift, *Americans Rely Less on Credit Cards than in Previous Years*, GALLUP.COM (Apr. 25, 2014), <http://www.gallup.com/poll/168668/americans-rely-less-credit-cards-previous-years.aspx>.

⁵⁶ Scott D. Schuh & Joanna Stavins, *The 2011 and 2012 Surveys of Consumer Payment Choice*, FED. RESERVE BANK OF BOSTON RESEARCH PAPER SERIES RESEARCH DATA REPORTS (2014), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2564165.

⁵⁷ 15 U.S.C. § 1643(a)(1)(B); 12 C.F.R. 226.12.

for card fraud.⁵⁸ Despite these policies, cardholders may still experience out-of-pocket losses if they do not report lost or stolen cards quickly enough.

According to the 2012 BJS survey, the average combined direct and indirect loss from existing-account credit card fraud was about \$1400 for the 69% of people who experienced any loss.⁵⁹ In 2012, it was about \$1000, with 66% experiencing a loss.⁶⁰ Based on these numbers, I estimate the range of average cost per existing-account credit card fraud at \$1,000 to \$1,400 with a point estimate of \$1200.

The 2012 and 2014 BJS surveys reported that victims of existing-account credit card fraud spent an average of 3 and 4 hours, respectively, resolving problems. A 2006 FTC report indicated that victims of existing-account credit card fraud spent a median of 2 hours resolving problems.⁶¹ I therefore use a range of 2–4 hours for t_k with a point estimate of 3.

For the cost of time parameter, I assume an average annual wage of \$45,500 per full-time employee, discounted 50% on the assumption that most time spent responding to breach occurs during non-work time.⁶² This corresponds to a \$12–\$20 cost of time, with a point estimate of \$15.

The estimate for f_k is dominated by the monetary cost of fraud, as shown in Table 3, with a range of about \$1,027 to \$1,505 with a point estimate of \$1,255.

Table 3: Expected cost per card of an existing-account credit card fraud incident

<i>Description</i>	<i>Low</i>	<i>Point</i>	<i>High</i>
Mean monetary cost of existing-account card fraud (c_{mk})	\$1,000	\$1,200	\$1,400
Mean hours spent responding to existing-account card fraud (t_k)	2	3	4
Cost of time per hour (c_{t_k})	\$12	\$15	\$20
Cost of reissuing cards used for fraud (c_{i_k})	\$3	\$10	\$25
Total expected cost of an existing-account card fraud incident (f_k)	\$1,027	\$1,255	\$1,505

2.5. Analysis

Table 1 summarizes the basic analysis of the per-card cost of reissuing versus not reissuing cards, with ranges and point estimates. The model estimates the expected cost of not reissuing cards at between \$1.15 and \$310 per card, with a point estimate of \$32.80. This wide range corresponds to a potential savings of about \$24 per card or loss of \$307 per card. The point estimate is a \$22.80 per-card loss by not reissuing. Multiplying these estimates by the number of reported breached card accounts implies that \$960 million might be lost by not reissuing cards immediately after a breach. The range of estimation is extreme, however: over

⁵⁸ Douglas Akers et al., *Overview of Recent Developments in the Credit Card Industry*, 17 FDIC BANKING REV. 3 (2005).

⁵⁹ HARRELL, *supra* note 50.

⁶⁰ HARRELL AND LANGTON, *supra* note 51.

⁶¹ SYNOVATE, *supra* note 52.

⁶² U.S. CENSUS, 2012 STATISTICAL ABSTRACT OF THE UNITED STATES t.647 (2012).

\$1 billion might be saved by not reissuing cards, but the potential total loss calculated by this model is almost \$14 billion.

Table 4: Comparison of the per-card cost of reissuing vs. not reissuing cards

<i>Description</i>	<i>Low</i>	<i>Point</i>	<i>High</i>
Reissue cost, per card	\$3.00	\$10.00	\$25.00
Expected cost if not reissued, per card	\$1.15	\$32.80	\$310.00
Per-card savings (cost) from not reissuing cards	(\$307.00)	(\$22.80)	\$23.85
Payment card records reported lost in data breaches per year (<i>nd</i>) (mil.)	39	42	45
Cumulative savings (cost) from not reissuing (mil.)	(\$13,800)	(\$960)	\$1,080

2.5.1. Monte Carlo Analysis

Monte Carlo simulations allow an estimation of the distribution of likelihood along the broad range of results. Because I have no reason to assume any particular distribution for the parameters, I used PERT Beta distributions with the highs, lows, and point estimates of the ranges as the equivalent values of the distributions. The resulting distributions show a wide variation in possible costs, with some overlap between the reissue and no-reissue situations.

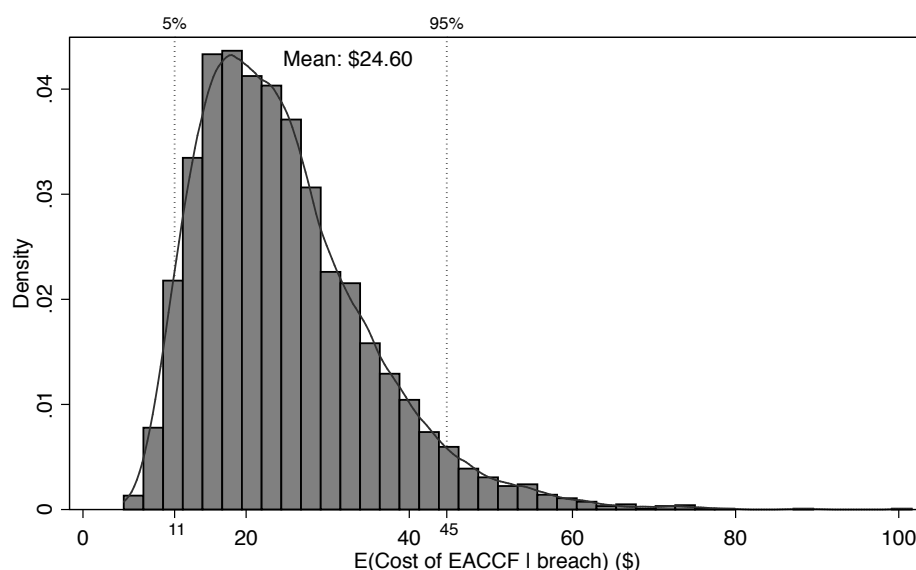


Figure 3: Distribution of the cost per card to reissue or not reissue cards based on a Monte Carlo simulation

Figure 3 shows a histogram of the expected per-card cost of fraud when cards are not reissued. The 90% confidence range is from \$11.20 per card to \$44.60 per card, with a mean of \$24.60. The distribution resembles the heavy-tailed models found for cyber-risk and data breach in previous work in the literature.⁶³ Figure 4 shows a histogram of the total cost reduction that could be achieved from not automatically reissuing credit cards. The 90% confidence range is

⁶³ See Edwards, Hofmeyr, and Forrest, *supra* note 14; T. Maillart & D. Sornette, *Heavy-Tailed Distribution of Cyber-Risks*, 75 EUR. PHYSICAL J. B 357 (2010).

(\$-1.4 billion, \$88 million), with about a 91% probability that immediately reissuing cards would be the lower-cost option.

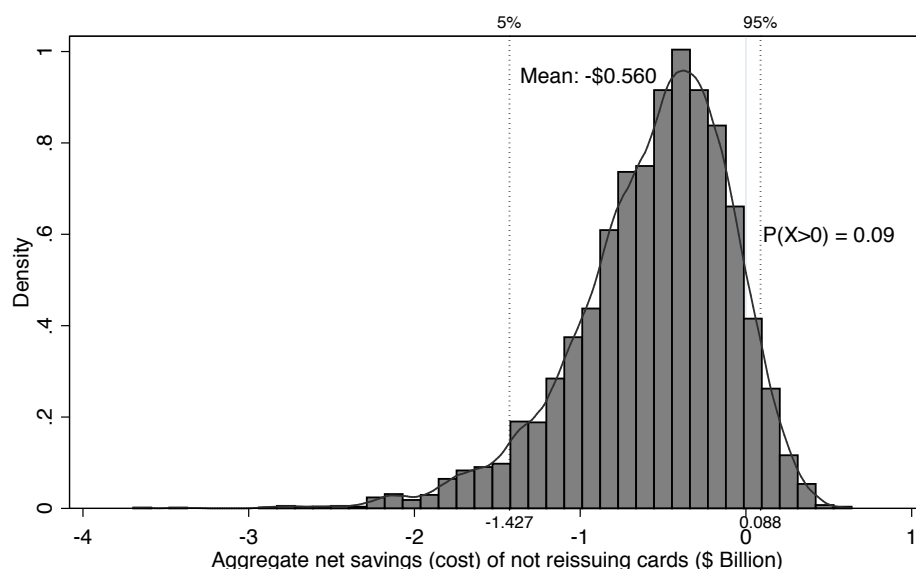


Figure 4: Histogram of cumulative savings from not automatically reissuing cards according to a Monte Carlo simulation

2.5.2. Sensitivity Analysis

Figure 5 is a tornado diagram showing the sensitivity of the per-card cost of not reissuing cards for the variables to which the cost is most sensitive. Each row shows the effect on the mean result of increasing the parameter by one standard deviation.

Unsurprisingly, the model is most sensitive to the parameters with the greatest uncertainty. The number of reported breaches with unknown record counts and the scaling factor for unreported breaches are both significant factors in the estimate. Each of these parameters reduce the mean estimate by over \$5 of the roughly \$25 mean expected cost of fraud from not reissuing. The model is also particularly sensitive to the percentage of existing-account credit card fraud attributable to breach. An increase in that parameter by one standard deviation increases the mean expected cost of fraud by about \$4.50. A fourth parameter that is not well-understood—the percentage of breached cards that are reissued—also has a large effect.

Although this chapter focuses on parameter uncertainty, the results can change dramatically due to model uncertainty. As mentioned in Section 4.2.6, BJS statistics on identity theft were originally collected by household, then more recently by individual. I use the individual-level data in the model because it avoids potential issues involving multiple cardholders per account in a household and reduces the number of instances in which one unit suffered multiple fraud incidents, violating one of the assumptions. When I began this work, however, individual-level data was not available and I used per-household data. The results of the model when I use per-household calculations are quite different than those initial results, even accounting for other refinements to the model since my initial work.

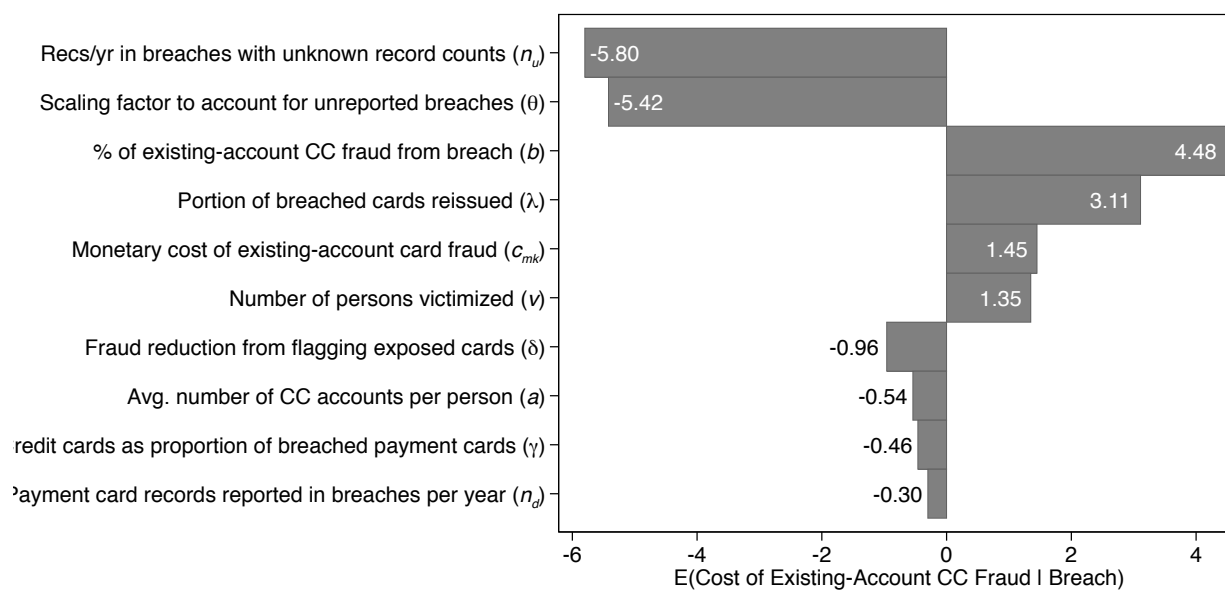


Figure 5: Tornado diagram of variables affecting the per-card cost of not reissuing cards

The only parameters that change are the number of cards per household and the number of households victimized by existing-account identity fraud. The number of cards per household is roughly similar to the number of cards per person—a range of 3.1 to 5.8 depending on year and source, as calculated by total cards divided by number of households—but the number of households experiencing existing-account credit card fraud was between 3.6 million and 5 million according to the BJS 2005-2010 survey.⁶⁴ As a result, the range of a per-household calculation would be an expected cost of credit card fraud on breached cards of between \$0.42 and \$170 per card, with a point estimate of about \$14—about \$18 less than the estimate using an individual-level calculation.

2.6. Discussion and Limitations

In answer to the question posed in the title of this chapter, reissuing cards immediately after a breach appears to be less costly than waiting for attempted fraud before reissuing. This result is fairly robust despite the wide uncertainty in the estimated cost of fraud after a breach. The Monte Carlo analysis estimates a 9% probability that waiting to reissue cards until fraud is detected would save money. The uncertainty in the model is partly because I rely on public data sources for the parameters and partly because the data sources themselves are subject to tremendous uncertainty.

The sensitivity analysis suggests where resources could be best targeted to getting better data for parameters critical to the 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

⁶⁴ LYNN LANGTON, U.S. DEPT. OF JUSTICE, IDENTITY THEFT REPORTED BY HOUSEHOLDS, 2005-2010 (2011), <http://www.bjs.gov/index.cfm?ty=pbdetail&iid=2207>.

inadequate; too many people simply do not know how their data 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 this data would undoubtedly improve our understanding of the benefits of options following a data breach. Access to that data would come with its own costs, of course, whether through compliance with a regulatory data-sharing regime or through costs of voluntary industry data-sharing. A comparison of the costs and benefits of increased data sharing by card issuers would be an opportunity for future work.

A data reporting regime may create its own perverse incentives. Participants in the card ecosystem who have full knowledge of the model used to make policy decisions might have incentives to manipulate that data. This incentive effect of disclosure is another topic worthy of future study.

My work building a database of credit card breaches shows that despite extensive breach reporting requirements, information about breaches is often incomplete. More states could follow lead of Maine, Maryland, and New Hampshire by requiring not only that breached organizations report the breach to the state attorneys general but provide detailed information about the breach, such as the number of residents affected, the cause of the breach, and the type of data breached. If states could agree on a standard form for breach reporting, the burden on reporting organizations could be held to a minimum.

2.6.1. Limitations and Opportunities for Future Work

The analysis in this chapter is subject to several limitations, each of which presents an opportunity for future study. One obvious and major limitation (as well as motivation) of this work is the lack of data on the causes, extent, and effects of data breach. Efforts such as the National Cyber Leap Year have attempted to fill this gap,⁶⁵ but much more work is needed to create the type of data that can be used for reliable statistical analysis. The implications of the poor quality of available data are discussed more fully in previous work.⁶⁶

Another limitation of the analysis described in this chapter is that it treats breaches as homogeneous—assuming, for example, that a small number of records in an improperly discarded report creates the same risk of data exploitation as the hacking of a large database. In particular, the model used in this analysis takes limited account of the wide variation in breach size. It might, for example, be social optimal to reissue cards after “everyday” breaches but not after megabreaches of one million cards or more, or it might be worth reissuing after hacking breaches but not after breaches due to improperly discarded records.

⁶⁵ Fred Chong et al., *National Cyber Leap Year Summit 2009: Co-chairs’ Report* (2009), https://www.nitrd.gov/nitrdgroups/index.php?title=National_Cyber_Leap_Year_Summit_2009 (last visited Nov 7, 2015).

⁶⁶ James T. Graves, Alessandro Acquisti & Nicholas Christin, *Big Data and Bad Data: On the Sensitivity of Security Policy to Imperfect Information*, 83 CHICAGO L. REV. 117 (2016).

This analysis does not account for “second-order” effects—those indirect costs that occur over time or that are in some other sense a step removed from the immediate costs. For example, immediately reissuing cards reduces the window during which thieves can attempt fraud, which can both dissuade credit card theft and make attribution and detection of fraud easier. Another second-order effect lies in cardholder behavior after his or her card has been affected in a breach. Cardholders may expect to have cards reissued automatically and reduce card usage—and perhaps overall spending—if they are not. These second-order costs weigh in favor of reissuing, thus strengthening the case for immediate reissue of breached cards.

This work is limited by lack of access to transaction-level card data. A researcher with industry access could improve on this work by combining the analysis of public data I present here with data on issuers’ costs. Data held by issuers could yield information about fraud probability and losses by cardholder demographics, breach attributes, and so forth.

2.6.2. Conclusion

Having determined that immediately reissuing cards appears to have a lower social cost, what are the policy implications? As mentioned at the beginning of this chapter, card association rules allow issuers to recover fraud costs that result from breached cards but not the operational costs of reissuing them. The card association rules may create incentives for issuers to wait before reissuing cards, which is the opposite of what the model suggests to be the socially optimal incentive. Limited evidence suggests that card issuers often do routinely re-issue cards affected in a breach despite these incentives.⁶⁷ If in fact this practice is widely followed, this research suggests that it is socially optimal.

⁶⁷ See *supra* Section 2.4.2.4.

3. Perception Versus Punishment in Cybercrime⁶⁸

3.1. Introduction

The U.S. Computer Fraud and Abuse Act (CFAA)⁶⁹ is not a popular law.⁷⁰ Enacted in 1986 to deal with the nascent computer crimes of that era, it has aged badly. It has been widely criticized as vague, poorly structured, and having an overly broad definition of loss that invites prosecutorial abuse.⁷¹ These criticisms only increased when Aaron Swartz committed suicide in 2013 after he was threatened with up to 35 years in prison for downloading millions of academic papers from an online database.⁷²

One of the problems with sentencing under the CFAA has received little attention: a misalignment between the facts that affect sentencing and the importance of those facts to the seriousness of CFAA crimes. It has been observed, for example, that CFAA sentences escalate

⁶⁸ This chapter was previously published as James T. Graves, Alessandro Acquisti, and Ross Anderson, *Perception Versus Punishment in Cybercrime*, 109 J. CRIM. L. & CRIMINOLOGY 313 (2019).

⁶⁹ 18 U.S.C. § 1030 (2018).

⁷⁰ See, e.g., Grant Burningham, *The Most Hated Law on the Internet and Its Many Problems*, NEWSWEEK (Apr. 16, 2016), <http://www.newsweek.com/most-hated-law-internet-and-its-many-problems-cfaa-448567> (describing criticisms of the CFAA by defense attorneys and security researchers); Brian Feldman, *Our Legal System Has No Idea How to Handle Computer Crimes*, N.Y. MAGAZINE (Apr. 14, 2016), <http://nymag.com/selectall/2016/04/matthew-keys-sentencing-computer-crimes.html> (describing the CFAA as “lagging 30 years behind” technology and “pos[ing] a danger to anyone who touches a computer”); Molly Sauter, *Online Activism and Why the Computer Fraud and Abuse Act Must Die*, BOING BOING (Sept. 26, 2014), <https://boingboing.net/2014/09/26/fuckthecfaa.html> (arguing that the CFAA criminalizes online activism).

⁷¹ See, e.g., Jennifer S. Granick, *Faking It: Calculating Loss in Computer Crime Sentencing*, 2 I/S: J. L. & POL’Y INFO. SOC’Y 207 (2006); Orin S. Kerr, *Vagueness Challenges to the Computer Fraud and Abuse Act*, 94 MINN. L. REV. 1561, 1561 (2010); Orin S. Kerr, *Cybercrime’s Scope: Interpreting “Access” and “Authorization” in Computer Misuse Statutes*, 78 N.Y.U. L. REV. 1596, 1616 (2003). See generally Paul J. Larkin, Jr., *United States v. Nosal: Rebooting the Computer Fraud and Abuse Act*, 8 SETON HALL CIR. REV. 257 (2012) (writing that “neither the text of the [CFAA] nor the litigation conducted to date draws a clear line separating lawful from unlawful conduct”); Andrea M. Matwyshyn, *The Law of the Zebra*, 28 BERKELEY TECH. L. J. 155 (2013) (arguing that “courts overzealously sanction defendants with CFAA penalties in addition to contract remedies”); Vasileios Karagiannopoulos, *From Morris to Nosal: The History of Exceeding Authorization and the Need for a Change*, 30 J. MARSHALL J. INFO. TECH. & PRIVACY L. 465, 477 (2014) (arguing that the case law provides a “confusing mix of interpretations” of the CFAA in the employment law context).

⁷² See, e.g., David Thaw, *Criminalizing Hacking Not Dating: Reconstructing the CFAA Intent Requirement*, 103 J. CRIM. L. & CRIMINOLOGY 907, 910 (2013); John Dean, *Dealing With Aaron Swartz in the Nixonian Tradition: Overzealous Overcharging Leads to a Tragic Result*, JUSTIA (25 Jan. 2013), <https://verdict.justia.com/2013/01/25/dealing-with-aaron-swartz-in-the-nixonian-tradition> (arguing that Swartz killed himself because the Boston U.S. Attorney’s Office “was planning to forever ruin him over an apparent act of civil disobedience”); Jennifer Granick, *Towards Learning from Losing Aaron Swartz*, CTR. FOR INTERNET & SOC’Y (Jan. 4, 2013), <http://cyberlaw.stanford.edu/blog/2013/01/towards-learning-losing-aaron-swartz> (discussing, shortly after Aaron Swartz’s suicide, his case and the problem of “prosecutorial overreaching”); Marcia Hoffmann, *In the Wake of Aaron Swartz’s Death, Let’s Fix Draconian Computer Crime Law*, EFF (Jan. 14, 2013), <https://www.eff.org/deeplinks/2013/01/aaron-swartz-fix-draconian-computer-crime-law> (discussing “extremely problematic elements” of the CFAA that made it possible for the government to “throw[] the book at Aaron for accessing MIT’s network and downloading scholarly research”).

rapidly as (easily inflated) losses increase.⁷³ But this escalation may be rapid not only in an absolute sense, but in disproportion to other attributes of the crime. Other factors, such as the offender's motivation, the context of the crime, its scope, or the type of data affected, may play a larger role in the seriousness of a crime.

The purpose of this piece is to explore that potential misalignment between punishment and perceptions through a series of empirical experiments that measure public opinions about cybercrime. Experimental measurement of public opinion has been used to study crime seriousness since at least the 1960s.⁷⁴ Criminal law codifies social norms, which manifest as perceptions that can be empirically measured.⁷⁵ More generally, public opinion influences policymaking.⁷⁶ Criminal codes “reflect through the state legislature’s deliberations and actions some understanding, however dim and remote, of what ‘the public’ deems appropriate for the crimes in question.”⁷⁷ Although public perceptions of the criminal justice system are flawed,⁷⁸ these perceptions influence how crimes are defined, what punishments they carry, whether those punishments are believed to be fair, and how resources are allocated to enforcement.

This chapter reports on the results of two studies with over 2,600 respondents: (1) a series of six between-subjects experiments and (2) a factorial vignette survey experiment. I conducted these two types of studies to take advantage of the benefits of each methodology. The factorial vignette methodology has been used to investigate how different factors of a crime (such as the offender's race, income, and gender) affect perceptions of that crime.⁷⁹ The between-subjects methodology, in contrast, allows us to ask more questions about each vignette as well as tailor the specifics of each vignette to increase plausibility.

These results provide empirical support for arguments that CFAA sentencing is miscategorized in the federal sentencing guidelines. Although an attacker's motivation, the type of data affected, and the amount of loss are all statistically significant factors in perceived seriousness, the weight placed on financial loss in sentencing calculations is not reflected in public attitudes. Another factor in CFAA sentencing—the target of the crime—appears to have

⁷³ See, e.g., Granick, *supra* note 71, at 211.

⁷⁴ See, e.g., Michael O’Connell & Anthony Whelan, *Taking Wrongs Seriously*, 36 BRIT. J. CRIMINOLOGY 299, 299 (1996); Section 3.2.2, *infra*.

⁷⁵ See Paul H. Robinson & John M. Darley, *The Utility of Desert*, 91 NW. U. L. REV. 2, 456–58 (1997); Paul H. Robinson et al., *The Origins of Shared Intuitions of Justice*, 60 VAND. L. REV. 1633, 1635 (2007).

⁷⁶ See, e.g., Amy L. Anderson et al., *Residency Restrictions for Sex Offenders: Public Opinion on Appropriate Distances*, 26 CRIM. JUST. POL’Y REV. 262, 263–64 (2015); Eric P. Baumer & Kimberly H. Martin, *Social Organization, Collective Sentiment, and Legal Sanctions in Murder Cases*, 119 AM. J. SOC. 131, 132 (2013); Paul Burstein, *The Impact of Public Opinion on Public Policy: A Review and an Agenda*, 56 POL. RES. Q. 29, 29–30 (2003); Justin T. Pickett et al., *Public (Mis)Understanding of Crime Policy: The Effects of Criminal Justice Experience and Media Reliance*, 26 CRIM. JUST. POL’Y REV. 500, 501 (2015).

⁷⁷ Peter H. Rossi et al., *Beyond Crime Seriousness: Fitting the Punishment to the Crime*, 1 J. QUANTITATIVE CRIMINOLOGY 59, 60 (1985).

⁷⁸ See generally, e.g., Julian V. Roberts, *Public Opinion, Crime, and Criminal Justice*, 16 CRIME & JUST. 99 (1992) (noting that the public has limited knowledge of the criminal justice system, holds misperceptions about crime rates and other statistics, and may be biased by sensationalistic news coverage).

⁷⁹ See *infra* note 166 and accompanying text.

no statistically significant effect on perceptions. In contrast, the most important factor in ratings of seriousness—the attacker’s motivation—has much less of an effect on sentencing. These results suggest that CFAA sentences are indeed out of alignment with the public’s views.

The rest of this chapter proceeds as follows. Section 3.2 provides background information. In Section 3.2.1, I discuss the factors that affect the maximum sentences under the CFAA and the factors that determine the recommended sentences under the federal sentencing guidelines. In Section 3.2.2, I summarize previous work on crime seriousness. Section 3.3 presents the methodology, model, and results of the between-subjects experiments. Section 3.4 presents the factorial vignette survey experiment. Section 3.5 discusses the implications of the results and concludes.

3.2. Background

3.2.1. Factors Affecting Sentencing Under the Computer Fraud and Abuse Act

As with all non-capital federal crimes, sentencing under the CFAA is determined by statutory provisions and federal sentencing guidelines. The statute sets maximum sentences based on the nature of the crime.⁸⁰ The sentencing guidelines determine the recommended sentencing range based on aspects of both the crime and relevant conduct.⁸¹ The rest of this section discusses how various factors of a CFAA crime affect maximum and recommended sentences.

3.2.1.1. Maximum Sentences

The CFAA criminalizes six types of conduct as “computer crime.”⁸² In general terms, these are (1) obtaining information,⁸³ (2) accessing government computers,⁸⁴ (3) committing computer fraud,⁸⁵ (4) causing damage with or to a computer,⁸⁶ (5) trafficking in passwords,⁸⁷ and (6) extorting money by threatening to obtain information or damage a computer.⁸⁸ Table 5 summarizes the CFAA sections and the maximum sentences for each. As the table shows, the base maximum sentence for most CFAA crimes is one year except for computer fraud and

⁸⁰ See 18 U.S.C. § 1030(c).

⁸¹ See U.S. SENTENCING GUIDELINES MANUAL §§ 2B1.1, 2B2.3, 2M3.2, 2X1.1 (U.S. SENTENCING COMM’N 2018).

⁸² For in-depth discussions of the CFAA, *see generally* DEP’T OF JUSTICE, PROSECUTING COMPUTER CRIMES, <https://www.justice.gov/sites/default/files/criminal-ccips/legacy/2015/01/14/ccmanual.pdf>; Kerr, *Cybercrime’s Scope*, *supra* note 71.

⁸³ 18 U.S.C. § 1030(a)(1)–(2).

⁸⁴ *Id.* § 1030(a)(3).

⁸⁵ *Id.* § 1030(a)(4).

⁸⁶ *Id.* § 1030(a)(5).

⁸⁷ *Id.* § 1030(a)(6).

⁸⁸ *Id.* § 1030(a)(7).

extortion, which have maximum sentences of five years for a first offense,⁸⁹ and accessing national security information, with a maximum sentence of ten years for a first offense.⁹⁰

Table 5: CFAA Sections and Maximum Sentences

Section	Description	Max. Sentence
1030(a)(1)	Obtaining national security information	10 (20)
1030(a)(2)	Obtaining information	1 or 5 (10)
1030(a)(3)	Accessing government computers	1 or 5 (10)
1030(a)(4)	Computer fraud	5 (10)
1030(a)(5)(A)	Intentional damage	1, 10, 20, or life (20 or life)
1030(a)(5)(B)	Reckless damage	1 or 5 (10)
1030(a)(5)(C)	Negligent damage	1 (10)
1030(a)(6)	Trafficking in passwords	1 or 5 (10)
1030(a)(7)	Computer extortion	5 (10)

Note: Maximum sentences for a second offense are listed in parentheses.

Two provisions can increase the maximum sentence. The first applies to CFAA crimes of accessing information, accessing government computers, or trafficking in passwords. The maximum sentence for any of these offenses increases to five years if (i) “the offense was committed for purposes of commercial advantage or private financial gain,” (ii) the offense was committed “in furtherance of any criminal or tortious act in violation of the Constitution or laws of the United States or of any State,” or (iii) “the value of the information obtained exceeds \$5000.”⁹¹

The other provision is a two-dimensional scale that increases maximum sentences for computer damage based on the amount of damage and the level of intent. Recklessly causing damage carries a maximum sentence of five years if the conduct led to at least \$5,000 in loss, impaired medical treatment, caused physical injury, posed a threat to public health or safety, damaged any computer used by the U.S. government “in furtherance of the administration of justice, national defense, or national security,” or damaged ten or more computers.⁹² If the offender intentionally caused any of the forms of damage listed above, the maximum sentence increases to ten years.⁹³ And if the offender intentionally caused serious bodily injury or death, the maximum sentence increases to twenty years or life, respectively.⁹⁴

If the data obtained in a cybercrime includes “a means of identification of another person,” the crime can be charged under the identity theft statutes.⁹⁵ A conviction for identity

⁸⁹ *Id.* § 1030(c).

⁹⁰ *Id.* § 1030(a).

⁹¹ *Id.* § 1030(a)(2), (c)(2)(B).

⁹² *Id.* § 1030(c)(4)(A)(i).

⁹³ *Id.* § 1030(c)(4)(B)(ii).

⁹⁴ *Id.* § 1030(c)(4)(E)–(F).

⁹⁵ *Id.* § 1028(a)(7). The offender must also have acted “with the intent to commit, or to aid or abet, or in connection with, any unlawful activity that constitutes a violation of Federal law, or that constitutes a felony under any applicable State or local law.”

theft carries a maximum sentence of five years.⁹⁶ Most computer-connected identity theft crimes will also subject the offender to prosecution under the aggravated identity theft statute, which adds two years imprisonment to a felony conviction under the CFAA.⁹⁷

Maximum sentences under the statute thus depend on the facts of a crime. The maximum sentence can increase based on scope, motive, consequences, context, and the type of information accessed. *Scope* refers to the number of victims. A CFAA crime that damages ten or more computers has a five-year maximum sentence based on scope.⁹⁸ *Motive* is reflected in an increased maximum sentence of five years for obtaining information for purposes of commercial advantage or financial gain.⁹⁹ The *consequences* of a CFAA crime can increase sentences through the \$5000 loss threshold in certain subsections¹⁰⁰ and through maximum sentences that grow longer as damage increases to include physical injury, serious bodily injury, or death.¹⁰¹ By *context*, I mean the type of organization or computer victimized. The increase in maximum sentence by five or ten years for damaging government computers is an example.¹⁰² And the *type of information* matters too: accessing identifying information such as social security numbers can increase the maximum sentence to five years or add two years to the imposed sentence.¹⁰³ If an offender accessed classified national security information, the maximum sentence for a first offense increases to ten years.¹⁰⁴

3.2.1.2. Sentencing Guidelines

Although the statute sets maximum sentences, sentence lengths within those maximums are largely determined by the federal sentencing guidelines. Promulgated by the United States Sentencing Commission pursuant to the Sentencing Reform Act of 1984,¹⁰⁵ the guidelines are intended to “provide certainty and fairness in meeting the purposes of sentencing, avoiding unwarranted sentencing disparities among defendants with similar records who have been found guilty of similar criminal conduct while maintaining sufficient flexibility to permit individualized sentences when warranted[.]”¹⁰⁶

The sentencing range recommended under the guidelines is a function of the crime’s offense level and the offender’s criminal history. To find the sentencing range for a particular

⁹⁶ *Id.* § 1028(b)(2)(B).

⁹⁷ *Id.* § 1028A(a)(1).

⁹⁸ *Id.* § 1030(c)(4)(A)(i), (B)(i).

⁹⁹ *Id.* § 1030(c)(2)(B).

¹⁰⁰ *Id.* § 1030(a)(4), (c)(2)(B), (c)(4)(A)(i)(I), (c)(4)(B)(i).

¹⁰¹ *Id.* § 1030(c)(4)(A)(i)(III), (c)(4)(B)(i), (c)(4)(E), (c)(4)(F).

¹⁰² *See id.* § 1030(a)(5), (c)(4)(A)(i)(V), (c)(4)(B)(i).

¹⁰³ *Id.* §§ 1028(b)(2)(B), 1028A(a)(1).

¹⁰⁴ *Id.* § 1030(a)(1), (c)(1).

¹⁰⁵ Sentencing Reform Act of 1984, Pub. L. 98-473, 98 Stat. 1987 (codified as amended at 18 U.S.C. §§ 3511–3673, 28 U.S.C. §§ 991–998).

¹⁰⁶ 28 U.S.C. § 991(b)(1)(B).

conviction, a court determines the offense level and criminal history category then consults the table reproduced in this chapter in Table 29. The offense level and criminal history category intersect at a sentencing range in months.

The offense level depends primarily on characteristics of the crime itself, such as the number of victims, amount of loss, and mitigating or aggravating factors, although offender characteristics can also play a part. For example, minimum offense levels apply to “career offenders.”¹⁰⁷ The criminal history category is based on the offender’s previous convictions and the length of previous sentences. Someone with no prior offenses has a criminal history category of I.

Most CFAA offenses are sentenced under section 2B1.1 of the guidelines, which covers theft, fraud, and similar economic crimes.¹⁰⁸ The exceptions are (a)(1) (obtaining national security information), which is sentenced under section 2M3.2, and (a)(3) (accessing government computers) and (a)(7) (extortion), which are sentenced under section 2B2.3.¹⁰⁹ The base offense level for most CFAA crimes is six.¹¹⁰ Computer extortion has a base offense level of eighteen, and unauthorized access to national security information carries a base offense level of thirty.¹¹¹

One of the largest factors that can increase an offense level is the amount of loss caused. Section 2B1.1(b)(1) lists a sliding scale of enhancements based on the actual or intended loss resulting from the crime. As of the 2016 guidelines, the enhancements range from two levels for a crime with at least \$6,500 in loss to thirty levels for a crime with at least \$550 million in loss.¹¹² That increase is roughly equivalent to an additional 8 to 10 years in prison (although maximum sentences may reduce that difference). \$550 million may seem unlikely for a hacking crime, but the CFAA is prone to inflated loss calculations.¹¹³ For example, Aaron Swartz allegedly downloaded 4.8 million articles that cost \$19 each to download from JSTOR.¹¹⁴ Had his case gone to trial, prosecutors might have argued that JSTOR suffered \$90 million in losses.

The guidelines also prescribe harsher sentences for crimes with greater scope. For example, the 2015 guidelines provide for a two-level enhancement—roughly a 25% increase in

¹⁰⁷ U.S. SENTENCING GUIDELINES MANUAL § 4B1.1 (U.S. SENTENCING COMM’N 2018).

¹⁰⁸ *Id.* app. A (indexing statutes to sentencing guidelines sections).

¹⁰⁹ *Id.*

¹¹⁰ *Id.* § 2B1.1. Access to government computers that does not lead to obtaining national security information has a base offense level of four, *see* § 2B2.3, but because a two-point enhancement mirrors the language of 18 U.S.C. § 1030(a)(3) the effective base level is six.

¹¹¹ *Id.* §§ 2B3.2, 2M3.2.

¹¹² *Id.* § 2B1.1(b)(1). Section 2B2.3, which applies to access to a government computer, also uses this loss scale.

¹¹³ *See, e.g.,* Granick, *supra* note 71, at 214–18 (arguing that “the most easily measurable type of harm that accrues from a computer attack is both unrelated to the severity of the intrusion and subject to manipulation by victims”); Orin S. Kerr, *Trespass, Not Fraud: The Need for New Sentencing Guidelines in CFAA Cases*, 84 GEO. WASH. L. REV. 1544, 1556–58 (2016) (noting that losses in CFAA sentencing “are unpredictable and usually outside the defendant’s control.”).

¹¹⁴ Indictment, *United States v. Swartz*, No. 1:11-cr-10260 at 9 (D. Mass. July 14, 2011); Open Access à la Pirate Bay, SCIENCEGUIDE (JULY 26, 2011), <https://www.scienceguide.nl/2011/07/open-access-a-la-pirate-bay/> (last visited Dec. 14, 2016).

sentence length—for a crime with ten or more victims or at least one victim who suffered “substantial financial hardship.”¹¹⁵ If more than five victims suffered substantial financial hardship, the enhancement is four levels, while more than twenty-five victims suffering substantial financial hardship triggers a six-point enhancement.¹¹⁶

The picture that emerges is that the guidelines place tremendous importance on loss. A crime that caused substantial financial hardship to twenty-five or more victims receives a six-level enhancement—the same as \$40,000 in losses. But it is complicated. The enhancements for loss and number of victims are not independent because a computer crime with more victims may also be more costly.

The type of information obtained is another salient feature in the calculation. Enhancements include a two-point increase in offense level (with a minimum offense level of 12) when the crime involved the use or transfer of an “authentication feature” or “means of identification”¹¹⁷ and a separate two-point increase if the offense involved “an intent to obtain personal information” or “unauthorized public dissemination of personal information.”¹¹⁸ The penalty for accessing national defense information increases the base offense level from thirty to thirty-five if the information was classified Top Secret.¹¹⁹

Enhancements may also be based on the target of a crime (what I refer to as the “context”). If a CFAA crime involved a system used in critical infrastructure or “by or for a government entity in furtherance of the administration of justice, national defense, or national security,” the offense level increases by two.¹²⁰ An additional six-point enhancement applies if the offense caused “substantial disruption of a critical infrastructure.”¹²¹

These are only some of the provisions that can affect the calculation of offense level. Other adjustments could apply depending on the offender’s role in the crime,¹²² acceptance of

¹¹⁵ U.S. SENTENCING GUIDELINES MANUAL § 2B1.1(b)(2)(A). “Substantial financial hardship” includes, among other things, becoming insolvent, filing for bankruptcy, suffering “substantial loss” of a savings fund, and suffering “substantial harm” to the victim’s ability to obtain credit. *Id.* § 2B1.1, cmt.4(F).

¹¹⁶ *Id.* § 2B1.1(b)(2)(B)–(C). Prior to the 2015 amendments, there was no requirement for “substantial financial hardship.” A crime involving 10 or more victims would receive a two-level enhancement, a crime involving 50 or more victims would receive a four-level enhancement, and a crime involving at least 250 victims would receive a six-point enhancement. *Id.* § 2B1.1(b)(2). The addition of “substantial financial hardship” to the criteria suggests that the sentencing commission wanted to de-emphasize the effect of scope.

¹¹⁷ *Id.* § 2B1.1(b)(11).

¹¹⁸ *Id.* § 2B1.1(b)(17).

¹¹⁹ *Id.* § 2M3.2.

¹²⁰ *Id.* § 2B1.1(b)(18)(i). Section 2B2.3 of the guidelines, applying to trespass, contains a similar provision.

¹²¹ *Id.* § 2B1.1(b)(18)(iii).

¹²² *See id.* §§ 3B1.1–3B1.5.

responsibility,¹²³ use of a “special skill”¹²⁴ or “sophisticated means,”¹²⁵ and motivation.¹²⁶ Many of these may easily apply to certain crime patterns. For example, damage to government computers for political purposes might qualify for enhancement based on “terrorism” as a motive.¹²⁷

3.2.2. Criminological Studies of Crime Seriousness

Criminologists have been studying perceptions of crime seriousness for nearly a hundred years.¹²⁸ In 1922, Willis Clark asked 100 people to “grade” on a scale from one to ten the seriousness of 148 acts of delinquency committed by schoolboys.¹²⁹ Categorizing these acts into different types (truancy, stealing, “incurability,” “malicious mischief,” etc., up to and including murder), Clark generated a numerical valuation for the seriousness of each offense.

Despite Clark’s work and other early efforts,¹³⁰ Sellin and Wolfgang are generally credited with pioneering empirical research.¹³¹ They sought to create a data-based index of delinquency that could be used to evaluate the effectiveness of efforts to combat juvenile crime.¹³² Although much of their work involved measuring and classifying delinquency based on statistics such as offense rates, they also believed that a measure of delinquency must account for seriousness.¹³³ They therefore conducted the first rigorous and comprehensive empirical study of attitudes towards crime, surveying judges, police, and college students in Philadelphia to

¹²³ See *id.* §3E1.1.

¹²⁴ *Id.* § 3B1.3 (U.S. Sentencing Comm’n 2016). A “special skill” is defined as “a skill not possessed by members of the general public and usually requiring substantial education, training or licensing.” *Id.* § 3B1.3 cmt.4.

¹²⁵ *Id.* § 2B1.1(b)(10)(C). Sophisticated means are defined as “especially complex or especially intricate offense conduct pertaining to the execution or concealment of an offense.” *Id.* § 2B1.1 cmt.9(B). Unlike the special-skills enhancement, which applies to all crimes, the sophisticated-means enhancement applies only to calculations under section 2B1.1.

¹²⁶ See *id.* §§ 3A1.1, 3A1.4.

¹²⁷ See 18 U.S.C. § 2332b(g)(5) (2015); U.S. SENTENCING GUIDELINES MANUAL § 3A1.4.

¹²⁸ For comprehensive reviews of the crime seriousness literature, see generally Gary Sweeten, *Scaling Criminal Offending*, 28 J. QUANTITATIVE CRIMINOLOGY 533, 533 (2012) (reviewing “a century of research on creating theoretically meaningful and empirically useful scales of criminal offending”); Stelios Stylianou, *Measuring Crime Seriousness Perceptions: What Have We Learned and What Else Do We Want to Know*, 31 J. CRIM. JUST. 37 (2003) (reviewing empirical studies of crime seriousness perceptions from 1964 through 2000).

¹²⁹ Willis W. Clark, CAL. BUREAU OF JUV. RES. BULL. 11, WHITTIER SCALE FOR GRADING JUVENILE OFFENSES (1922); see also John Henderson Gorsuch, *Scale of Seriousness of Crimes*, 29 J. CRIM. L. & CRIMINOLOGY 245, 245 (1938).

¹³⁰ See Sweeten, *supra* note 128, at 535–37.

¹³¹ See, e.g., Peter H. Rossi et al., *The Seriousness of Crimes: Normative Structure and Individual Differences*, 39 AM. SOC. REV. 224, 225 (1974) (“The most extensive previous treatment measuring crime seriousness is the pioneering work of Sellin and Wolfgang”); Stylianou, *supra* note 128, at 37 (“The study of perceptions of crime seriousness was introduced by Sellin and Wolfgang”).

¹³² THORSTEN SELLIN & MARVIN E. WOLFGANG, *THE MEASUREMENT OF DELINQUENCY* 1 (1964).

¹³³ *Id.* at 6.

come up with rankings for 141 different offenses.¹³⁴ Other scholars soon replicated and extended their work.¹³⁵

In the half century since then, the study of crime seriousness has continued to be an active area of criminological research. The threads developed in that area of research tackle different questions: What is “seriousness?” What are its components? What are the properties of a useful seriousness scale? How do people form judgments of seriousness? By what methodologies can it be measured? Is there a consensus on the seriousness of crimes? What are the perceptions of crime seriousness?

The first of these questions is fundamental—if we do not know what we mean by seriousness, how can we expect to measure it? We could define it as a partial order on punishment: one crime is more serious than another if and only if it should be punished more harshly. Some hope for an additive property, such that a crime that is twice as serious as another should receive twice as harsh a penalty. This question of additivity is a significant issue. Sellin and Wolfgang’s effort to create an additive scale is one of the reasons their work is considered seminal.

Several researchers have studied the components or dimensions of seriousness. Mark Warr identified two dimensions: the moral wrongfulness of the crime and the harmfulness of the offense’s consequences.¹³⁶ He asked Dallas residents to rate the seriousness, wrongfulness, and harmfulness of 31 crimes. His results were mixed. Among some respondents, different dimensions predominated for different classes of crimes (e.g., property crimes versus public order crimes) and wrongfulness and harmfulness were good predictors of seriousness.¹³⁷ Other respondents appeared to ignore moral wrongfulness entirely, judging crimes solely on the harm done.¹³⁸

Warr’s decomposition was relatively simple. Others have proposed more dimensions. Mark Hansel, for example, analyzed seriousness along nine dimensions: actual harm, potential harm, harmfulness to the offender, the “sickness” of the offense, the extent to which the offense is “personal,” and whether the offense is property related, violent, immoral, or sex-related.¹³⁹

¹³⁴ *Id.* at 241–58.

¹³⁵ See generally, e.g., Monica A. Walker, *Measuring the Seriousness of Crimes*, 18 BRIT. J. CRIMINOLOGY 348 (1978) (extending Sellin & Wolfgang’s work to a general population sample and confirming consistency of results across multiple methods); Peter H. Rossi et al., *supra* note 131, at 224 (surveying households in Baltimore to obtain ratings of a set of 140 crimes).

¹³⁶ Mark Warr, *What is the Perceived Seriousness of Crimes?*, 27 CRIMINOLOGY 795, 796 (1989). Sean Rosenmerkel replicated this work several years later, focusing on white-collar crimes. See Sean Rosenmerkel, *Wrongfulness and Harmfulness as Components of Seriousness of White-Collar Offenses*, 17 J. CONTEMP. CRIM. JUST. 308, 313 (2001).

¹³⁷ Warr, *supra* note 136, at 802–08.

¹³⁸ *Id.* at 810–15.

¹³⁹ Mark Hansel, *Citizen Crime Stereotypes—Normative Consensus Revisited*, 25 CRIMINOLOGY 455, 460 (1987).

Stephen Blum-West looked at eight dimensions: bodily harm, economic damage, emotional damage, potential for harm, intent, purpose, motive, and fair play.¹⁴⁰

Measurements of the components of seriousness naturally lead into questions of other factors that might affect perceptions. In contrast to studies such as Sellin and Wolfgang's, which attempt to rank a broad range of crimes, these studies are primarily concerned with how perceptions are affected by characteristics of the offenders, victims, and crime circumstances. Thus, while the Sellin and Wolfgang study and its direct progeny asked respondents to rate a relatively large number of short and general crime descriptions, studies of crime factors sometimes present fewer but longer and more detailed scenarios.

Although some crime factor studies have presented respondents with a single scenario¹⁴¹—and indeed I use a similar approach in one of the studies—it is also common to ask respondents to rate multiple scenarios. One technique is the factorial vignette survey experiment, which has been used to study normative and positive judgments.¹⁴² In this kind of experiment, respondents rate a series of short paragraph-length vignettes. Each describes the same basic scenario, but with different details. For example, a study of perceptions of just punishments for street crimes might use a template describing a robbery; each vignette would describe a version that differs in details such as the offender's and victim's age, race, gender, and whether a dangerous weapon was used. If the values (or “levels”) for each of the variables (“factors” or “dimensions”) are randomly generated, the factorial survey has many of the features of a fully randomized experiment—a regression analysis based on ordinary least squares (OLS) is expected to generate unbiased coefficients.¹⁴³ And although the total number of combinations of factors and levels (the “vignette space”) may be very large, the response set is also large because each respondent rates several vignettes.¹⁴⁴

Rossi, Simpson, and Miller were among the first to apply the factorial vignette methodology to perceptions of crime seriousness.¹⁴⁵ They presented 774 respondents with 50 vignettes describing a crime for which a person had been convicted. The vignettes varied over 20 dimensions, including 57 crime descriptions, 7 amounts of money stolen, 4 degrees of previous violations, 8 ranges for the age of the offender, and so on. They used a computer

¹⁴⁰ Stephen Blum-West, *The Seriousness of Crime: A Study of Popular Morality*, 6 *DEVIANT BEHAV.* 83 (1985).

¹⁴¹ See, e.g., Mary Dodge et al., *Do Men and Women Perceive White-Collar and Street Crime Differently? Exploring Gender Differences in the Perception of Seriousness, Motives, and Punishment*, 29 *J. CONTEMP. CRIM. JUST.* 399, 403 (2013).

¹⁴² See KATRIN AUSPURG & THOMAS HINZ, *FACTORIAL SURVEY EXPERIMENTS* 13–15 (2015); Guillermina Jasso, *Factorial Methods for Studying Beliefs and Judgments*, 34 *SOC. METHODS & RES.* 334, at 338–39; Rossi et al., *Beyond Crime Seriousness: Fitting the Punishment to the Crime*, 1 *J. QUANTITATIVE CRIMINOLOGY* 59, 62.

¹⁴³ See Rossi et al., *supra* note 77, at 68–69.

¹⁴⁴ See *id.* For example, Rossi, Simpson, and Miller's 1985 study used 20 dimensions with 3 to 57 levels each for a vignette space of over one trillion unique vignettes (experts in factorial vignette methodology would almost certainly say today that 20 dimensions is far too many to expect respondents to keep track of). But because 774 respondents rated 50 vignettes each, Rossi and his colleagues had over 53,000 vignette ratings in their answer set—more than enough to estimate coefficients for each individual dimension.

¹⁴⁵ *Id.* at 62.

program to print booklets of 50 vignettes each that respondents rated on paper. The rating task was to mark an unnumbered line answering whether “The sentence given was . . .” with anchors for “much too low,” “low,” “about right,” “high,” and “much too high.” Their analysis showed that perceptions of a crime are affected by characteristics of the crime, its consequences, the offender, and the people making the judgments.

One of the questions raised by research into seriousness is the extent to which people agree in their judgments. Blumstein and Cohen studied consensus in a 1980 study.¹⁴⁶ They asked residents of western Pennsylvania to assign sentences to 23 crimes and compared their recommendations to actual sentences. Respondents tended to agree on the relative severity of crimes but disagreed over the appropriate magnitude of punishment. They also tended to recommend more severe punishments than those actually imposed by courts. Rossi, Simpson, and Miller tackled consensus in their paper,¹⁴⁷ and Guillermina Jasso discusses it in depth in the context of measuring judgments using factorial vignette surveys.¹⁴⁸

Other work in studying crime seriousness has focused on particular types of crime. For example, criminologists have studied perceptions of white-collar crimes,¹⁴⁹ environmental crimes,¹⁵⁰ and “small” crimes.¹⁵¹ White-collar crimes are generally seen as less serious but their perceived seriousness appears to have increased over the years since Wolfgang and Sellin’s 1964 study.¹⁵² Although white-collar crimes may be similar to computer crimes, to my knowledge there has been only one published study of attitudes about cybercrime,¹⁵³ and none that has analyzed how the features of cybercrimes affect perceptions.

3.3. Study I: Between-Subjects Experiments

To understand how features of cybercrimes affect individuals’ perceptions, I conducted two human-subjects studies whose methodologies complement each other (see Section 3.1).

¹⁴⁶ Alfred Blumstein & Jacqueline Cohen, *Sentencing of Convicted Offenders: An Analysis of the Public’s View*, 14 L. & SOC. REV. 223, 248–52 (1980).

¹⁴⁷ Rossi et al., *supra* note 77, at 81–89.

¹⁴⁸ See Jasso, *supra* note 142, at 388–403.

¹⁴⁹ See generally, e.g., Francis T. Cullen et al., *The Seriousness of Crime Revisited: Have Attitudes Toward White-Collar Crime Changed?*, 20 CRIMINOLOGY 83 (1982) (studying whether perceptions of white-collar crime had changed since 1972 more than perceptions of other kinds of crime); Dodge et al., *supra* note 73 (studying perceptions of white-collar crimes versus street crimes with a focus on gender); Sean Rosenmerkel, *Wrongfulness and Harmfulness as Components of Seriousness of White-Collar Offenses*, 17 J. CONTEMP. CRIM. JUST. 308 (2001) (studying perceptions of white-collar offenses as compared to property offenses and violent offenses).

¹⁵⁰ See generally Tara O’Connor Shelley et al., *What About the Environment? Assessing the Perceived Seriousness of Environmental Crime*, 35 INT’L J. COMP. & APPLIED CRIM. JUST. 307 (2011) (studying whether the public perceives environmental crimes to be serious crimes).

¹⁵¹ See generally Salima Douhou et al., *The Perception of Small Crimes*, 27 EUR. J. POL. ECON. 749 (2011) (studying perceptions of “small crimes” such as littering, cheating on taxes, and speeding).

¹⁵² See Cullen et al., *supra* note 149, at 83, 92–94; Dodge et al., *supra* note 141, at 412 (2013).

¹⁵³ See Matthew B. Kugler, *Measuring Computer Use Norms*, 84 GEO. WASH. L. REV. 1568 (2016) (see discussion *infra* Section 4.2.2).

Study I consists of six between-subjects experiments and is discussed in this section. Study II is a factorial vignette survey experiment and is discussed in Section 3.4.

3.3.1. Methodology

3.3.1.1. Research Questions

I designed six between-subjects experiments, randomly assigning each subject to one experimental condition. In each experiment, I manipulated different features of a crime, one at a time. Each experiment relied on the presentation of a vignette describing an intentional data breach of consumers' personal information. I chose this for a number of reasons. The data breach scenario is a common one that I believe is readily understandable by most people.¹⁵⁴ It also lends itself to manipulation of the attributes of interest (scope, context, motivation, etc.) while holding other attributes reasonably constant.

The experiments focus on six aspects of cybercrime likely to influence perceptions of wrongfulness or harmfulness. Five of them are, as discussed in Section 3.2.1, directly relevant to sentencing: (1) scope, (2) motivation, (3) consequences, (4) context, and (5) the type of data affected. I also investigate (6) the breached organization's co-responsibility to learn whether people perceive a crime to be less serious when it was facilitated by an organization's poor security practices.

To study perceptions of these aspects, I use the following informal hypotheses:

- H1: Theft of medical data is seen as more wrongful and more harmful than the theft of name and address data.
- H2: Perceptions of crime harmfulness and severity increase with the number of records downloaded in a data breach.
- H3: A cybercrime committed by someone with a profit motive is seen as more wrongful than one committed by a political activist or a person curious about security vulnerabilities.
- H4a: A cybercrime with more expensive consequences is seen as more harmful, but not necessarily more wrongful, than cybercrimes causing less damage.
- H4b: People perceive cybercrimes as worse when large losses fall on consumers rather than on businesses.
- H5: An organization that had not patched its servers when it was breached is perceived as more co-responsible for the crime than an organization that had patched its servers.
- H6: Downloading data from a bank or government agency is perceived as more wrongful and harmful than downloading the same data from a non-profit.

¹⁵⁴ See Data Breaches, PRIVACY RIGHTS CLEARINGHOUSE, <http://www.privacyrights.org/data-breach> (last visited Aug. 31, 2018) (stating that over 11 billion data records have been affected in over 8,000 data breaches since 2005).

Each of the hypotheses listed above is *ceteris paribus*—that is, it is assumed that all factors not in the manipulation are held equal.

3.3.1.2. Design

Study I consisted of six between-subjects online survey experiments. Within each experiment, participants were randomly assigned to one of the conditions. Depending on the experiment, the number of conditions ranged from two to five. The six experiments manipulated the six aspects already discussed: type of data, scope, motivation, consequences, co-responsibility, and context.

One of the challenges was to manipulate only one attribute at a time. I was therefore careful to choose vignette language that minimized the possibility that a manipulation of one variable would “spill over” into an effect on consequences, which might dominate other manipulations. At the same time, vignettes had to be believable. I tackled these issues by specifying consequences whenever possible and by stating in the vignette that the perpetrator of the data breach in the scenario did not release the data he downloaded. This had the desirable side effect of limiting extreme “ceiling” effects in the responses to the questions. Because the consequences were minimalized, the answers in each vignette were better distributed across the range than they otherwise might have been.

All between-subjects experiments (and their conditions) followed the same structure. Participants who passed a screening process received an online survey. The survey asked them to read a vignette similar to the following:

On June 3, 2013, while browsing the Internet, Tom Smith discovered a security flaw in the Acme Insurance Company’s website. He used that flaw to gain access to Acme’s internal network and download 100,000 records from Acme’s customer database. Each record consisted of a customer’s full name, phone number, and address. Tom did not use or release the information. Acme’s customers suffered no harm.

Each experiment modified or extended this vignette with a particular manipulation. In the “Type of Data” experiment, the survey described the data obtained in the breach as either names, phone numbers, and addresses; or names, health history, medical diagnoses, and prescription records. The “Scope” experiment described the number of records downloaded as 10, 100, 1,000, 10,000, or 1,000,000 records depending on condition. In the “Motivation” experiment, the vignette included text explaining why Tom Smith was looking for security flaws—he was trying to make money, was a student looking to learn about computer security, or was an activist looking for evidence of corporate corruption. The “Consequences” experiment included three conditions: either Acme spent \$1000 to secure its servers, Acme spent \$5 million to repair damage to its database, or Acme’s customers suffered a collective \$5 million in identity theft. In the “Co-Responsibility” experiment, Acme had either patched its servers or not. In the

“Context” experiment, the organization from which Tom Smith downloaded the data was described as a bank, a non-profit organization, or a government agency.

After they read the vignette, participants saw a series of multiple-choice questions intended to test their recall of the details. Each experiment included questions to test recollection of the vignette’s data type, context, and scope. If these three questions did not include the manipulated variable, I added an additional question to check recall of the manipulation. After each memory-check question, the survey showed each participant a page indicating whether his or her answer was correct and repeating the correct answer to further reinforce the participant’s awareness of the details.

The survey then collected the variables of interest. Participants were asked to answer a series of questions on a 1–7 Likert scale. I selected the first three questions in accordance with previous research on the factors of crime seriousness.¹⁵⁵ The survey presented the following questions in random order:

- “How wrongful were Tom Smith’s actions?”
- “How serious was the crime Tom Smith committed?”
- “How harshly should Tom Smith be punished?”
- “How harmful were Tom Smith’s actions?”
- “How responsible was the Acme Insurance Company for the crime?”¹⁵⁶
- “How clever was Mr. Tom Smith?”
- “How sensitive were the data that Tom Smith downloaded?”

The survey also asked participants to recommend a specific punishment for the crime. The question was multiple-choice, with eleven options ranging from no punishment at all on the low end, to probation, to a sentence of 0–30 days, all the way to a sentence of life in prison on the high end, with intermediate sentence lengths in between.

In the Motivation, Consequences, Co-Responsibility, and Context experiments, the survey followed the specific-punishment question with a question about the potential consequences of Tom Smith’s actions. This question was intended to help determine whether participants judged scenarios by potential consequences instead of the actual consequences described in the scenarios. The added question also made another attention check possible: participants who rated the potential consequences as lower than the actual consequences may not have been paying enough attention to the questions. I removed these responses from the response set.

The next section included several questions intended to measure participants’ attitudes and experiences about data protection and personal privacy. I used the fifteen-question Concern

¹⁵⁵ See, e.g., Warr, *supra* note 136, at 796.

¹⁵⁶ In the Context experiment, the “Acme Insurance Company” was replaced by “ACR.”

for Information Privacy (CFIP) scale.¹⁵⁷ I also asked how often participants had suffered identity theft, how often they provide fake information when registering for web sites, and how much they had heard or read about “use and potential misuse of information collected from the Internet” in the past year. The survey instrument concluded with demographic questions and a few open-ended questions.

I ran ordered probit regressions on each variable of interest. Regressions included controls for demographics, memory check correctness, and privacy attitudes. I treated the demographic variables for gender, country of birth, age, education, occupation, work situation, and the memory check variables as categorical variables. I treated as continuous variables (1) the extent to which participants had been affected by cybercrime or privacy invasions and (2) the extents to which they use fake personal information and are aware of media coverage of data misuse.

3.3.2. Theoretical Model

For each experiment, I model a belief function of the form

$$Y = \beta_0 + \beta_1 X + \sum \gamma_q Z_q + \varepsilon \quad (5)$$

where each Y is a judgment about the crime, X is an attribute of that crime, $\gamma_q Z_q$ are attributes of the respondents q and their coefficients, and ε is the error term (which encompasses attributes of the crime other than X).

The model thus predicts a collective belief function with shared (or aggregate) intercept and slope. Although this is an overly simplistic model, it offers flexibility in evaluating multiple judgments.

3.3.3. Results

For each experiment, I used Amazon Mechanical Turk (MTurk) to recruit participants 18 years of age or older who lived in the United States, had at least a 95% approval rating on MTurk, and had not previously participated in any of the studies described in this chapter. The demographics and data quality of MTurk experiments have been extensively studied in multiple experimental contexts.¹⁵⁸ Several studies have shown that recruitment for online studies through

¹⁵⁷ See generally H. Jeff Smith et al., *Information Privacy: Measuring Individuals' Concerns About Organizational Practices*, 20 MIS Q. 167 (1996) (describing the development and test of an instrument for measuring individuals' levels of privacy concern).

¹⁵⁸ See generally, e.g., Michael Buhrmester et al., *Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality Data?*, 6 PERSP. PSYCH. SCI. 3, 3 (2011) (finding the data obtained with MTurk samples to be “at least as reliable as those obtained via traditional methods”); Matthew J.C. Crump et al., *Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research*, 8:3 PLOS ONE 1 (2013), <http://dx.doi.org/10.1371/journal.pone.0057410> (replicating several tasks from experimental psychology using MTurk and finding that most were “qualitatively successful”); Joseph K. Goodman et al., *Data Collection in a Flat*

MTurk can lead to more representative samples and better data quality than studies using other “convenience” samples such as university students.¹⁵⁹ Peer and his co-authors found that reputation alone is often enough to ensure sufficient data quality in MTurk studies.¹⁶⁰ Another study showed that MTurkers paid more attention to instructions than did traditional subject pool samples.¹⁶¹

The MTurk job description asked people to take “a short survey on crime.” I recruited a total of 2,635 participants in October through December 2013. I screened potential participants to exclude anyone who had participated in the crime seriousness experiments from participating in subsequent experiments in this series. I also filtered out responses with duplicated IP addresses or MTurk IDs, that claimed that the participant was under 18 years old or resided outside the U.S., or that contained contradictory answers rating the vignette’s potential consequences as greater than the actual consequences.

The remaining data set consists of 2,440 responses across six experiments. In each experiment the median age category is 25–34. Responses from females range from 41% to 52% of responses in each study. The only statistically significant difference across conditions in terms of age, gender, education, occupation, or work situation is (1) in the Motivation experiment, in which occupation differs at $p < 0.05$ and work situation differs at one-sided $p < 0.05$; and (2) the Context experiment, in which work situation differs between conditions at $p < 0.05$. I account for these variables (and all other demographic variables) in the regressions.

World: The Strengths and Weaknesses of Mechanical Turk Samples, 26 J. BEHAV. DECISION MAKING 213, 213 (2013) (finding that, despite “many similarities between MTurk participants and traditional samples,” MTurk participants could be less attentive and have lower self-esteem); Winter Mason & Siddharth Suri, *Conducting Behavioral Research on Amazon’s Mechanical Turk*, 44 BEHAV. RES. METHODS 1 (2012) (describing MTurk and discussing issues with MTurk research); Gabriele Paolacci et al., *Running Experiments on Amazon Mechanical Turk*, 5 JUDGMENT & DECISION MAKING 411 (2010) (reviewing MTurk and comparing it to other subject pools); Joel Ross et al., *Who Are the Crowdworkers? Shifting Demographics in Mechanical Turk*, in CHI ’10 EXTENDED ABSTRACTS HUM. FACTORS COMPUTING SYS. 2863 (2010) (describing how MTurk worker demographics have changed); Daniel J. Simons & Christopher F. Chabris, *Common (Mis)Beliefs about Memory: A Replication and Comparison of Telephone and Mechanical Turk Survey Methods*, 7:12 PLOS ONE 1 (Dec. 18, 2012), <http://dx.doi.org/10.1371/journal.pone.0051876> (using MTurk to replicate a telephone survey).

¹⁵⁹ See Tara S. Behrend et al., *The Viability of Crowdsourcing for Survey Research*, 43 BEHAV. RES. METHODS 800, 810–11 (2011); Adam J. Berinsky et al., *Evaluating Online Labor Markets for Experimental Research: Amazon’s Mechanical Turk*, 20 POL. ANALYSIS 351, 366 (2012) (concluding that “despite possible self-selection concerns, the MTurk subject pool is no worse than convenience samples used by other researchers in political science”); Krista Casler et al., *Separate but Equal? A Comparison of Participants and Data Gathered via Amazon’s MTurk, Social Media, and Face-to-Face Behavioral Testing*, 29 COMPUTERS HUM. BEHAV. 2156, 2158–59 (2013).

¹⁶⁰ Eyal Peer et al., *Reputation as a Sufficient Condition for Data Quality in Amazon Mechanical Turk*, 46 BEHAV. RES. METHODS 1023, 1030–31 (2014).

¹⁶¹ David J. Hauser & Norbert Schwarz, *Attentive Turkers: MTurk Participants Perform Better on Online Attention Checks than Do Subject Pool Participants*, 48 BEHAV. RES. METHODS 400, 405 (2016).

Table 6: Summary of Regression Results in Between-Subjects Experiments

Experiment & Condition	Wrongful	Harmful	Serious	Harshly	Sensitive	Respons.	Clever	Pot. Harmful	N
Type of Data									
Medical (v. Directory)	-0.104 (0.142)	0.194 (0.145)	0.076 (0.148)	-0.028 (0.145)	0.970*** (0.151)	0.015 (0.153)	0.008 (0.143)		239
Scope									
log(Records)	0.070** (0.27)	0.078** (0.026)	0.159*** (0.028)	0.107*** (0.026)	0.135*** (0.031)	0.064* (0.026)	0.057* (0.025)		583
Motivation									
Student (v. Profiteer)	-0.878*** (0.151)	-0.327* (0.148)	-0.596*** (0.150)	-0.793*** (0.145)	0.201 (0.141)	0.034 (0.141)	0.217 (0.141)	-0.051 (0.147)	361
Activist (v. Profiteer)	-0.795*** (0.150)	-0.279 (0.145)	-0.538*** (0.152)	-0.497*** (0.147)	0.130 (0.154)	0.100 (0.145)	0.191 (0.152)	-0.294 (0.159)	361
Consequences									
Acme (v. Low)	0.179 (0.123)	0.407*** (0.122)	0.083 (0.119)	0.338** (0.123)	0.147 (0.137)	-0.009 (0.140)	-0.123 (0.116)	-0.020 (0.118)	479
Customers (v. Low)	0.042 (0.125)	0.377** (0.120)	0.131 (0.121)	0.236* (0.118)	0.093 (0.138)	0.040 (0.151)	0.112 (0.126)	-0.125 (0.124)	479
Co-Responsibility									
Patched (v. Not)	0.133 (0.136)	0.102 (0.136)	0.157 (0.133)	0.074 (0.132)	0.087 (0.151)	-0.370* (0.164)	0.423*** (0.128)	-0.184 (0.136)	276
Context									
Gov't (v. Bank)	-0.055 (0.119)	0.013 (0.121)	-0.027 (0.125)	-0.030 (0.116)	0.147 (0.139)	-0.121 (0.142)	0.152 (0.118)	-0.023 (0.116)	502
Non-Profit (v. Bank)	0.048 (0.123)	-0.029 (0.124)	-0.222 (0.122)	0.030 (0.121)	0.099 (0.140)	-0.208 (0.155)	-0.361** (0.120)	-0.185 (0.121)	502

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6 shows the coefficients and standard errors for the eight variables of interest in all six experiments. The results of these experiments lead to the following conclusions for each of the hypotheses:

H1: Theft of medical data is seen as more wrongful and more harmful than the theft of name and address data.

As expected, participants rated names, health histories, medical diagnoses, and prescription records as more sensitive than names, phone numbers, and addresses ($p < 0.001$). The effect is strong as well as significant: 72% of participants in the medical-data condition rated the data as 7 (“Extremely sensitive”) or 6 compared to 34% of those in the directory-data condition.

Perceived crime severity, however, did not differ between conditions with statistical significance. Answers to “How sensitive was the data?” and “How serious was the crime?” are strongly correlated ($p < 0.001$, χ^2) but the difference in perceptions of data sensitivity by condition does not translate to a statistically significant difference in perceptions of crime severity.

H2: Perceptions of crime harmfulness and severity increase with the number of records downloaded in a data breach.

The number of records had a statistically significant effect in the expected direction on all the Likert-type question responses. Note, however, that this may be due in part to the large sample size compared to the other experiments. Although I kept the number of participants *per condition* about the same as in other experiments, the total number makes it more likely that small-magnitude results such as those seen for Acme's co-responsibility for the breach ($\hat{\beta} = 0.064$, $se = 0.026$, $p < 0.05$) and Tom's cleverness ($\hat{\beta} = 0.057$, $se = 0.025$, $p < 0.05$) will be statistically significant.

Interestingly, participants rated the data as more sensitive when more records were affected. The magnitude of that effect ($\hat{\beta} = 0.135$) is larger than that for any of the seven Likert questions except for seriousness ($\hat{\beta} = 0.159$). Interpreting this result is challenging without additional information, but two possible explanations seem plausible. First, the survey experiment may not have done an adequate job of asking about the sensitivity of the type of data downloaded as opposed to the sensitivity of the entire set of actual data records downloaded. Second, people may have conflated data sensitivity and the total potential for harm from the amount of data.

H3: A cybercrime committed by someone with a profit motive is seen as more wrongful than one committed by a political activist or a person curious about security vulnerabilities.

Participants judged the profiteer's crime as more serious than the same crime committed by a student or activist. There was virtually no statistically significant difference in perceptions of the student and the activist, however. Participants rated the profiteer's crime as more wrongful ($p < 0.001$), harmful ($p < 0.05$), and serious ($p < 0.001$) than the student's, and said that the crime should be punished more harshly ($p < 0.001$). The difference between the profiteer and activist was only slightly less pronounced, with strongly significant results for both wrongfulness ($p < 0.001$) and seriousness ($p < 0.001$), and with one-sided significance for harmfulness ($p < 0.05$). The profiteer also received harsher judgments, compared with the activist, of how harshly he should be punished ($p < 0.01$). And although participants said that the activist should be punished more harshly than the student ($p < 0.05$), perceptions of wrongfulness, harmfulness, and seriousness were statistically indistinguishable.

H4a: A cybercrime with more expensive consequences is seen as more harmful, but not necessarily more wrongful, than cybercrimes causing less damage.

The manipulation had the expected effect on perceptions of harmfulness. The conditions in which either Acme ($p < 0.001$) or its customers ($p < 0.01$) spent \$5 million received higher ratings of harmfulness than the condition in which the only cost was \$1,000 to secure servers

(the “Low” condition). Participants also said that each of these two cases should be punished more harshly than the Low condition (Acme: $p < 0.01$, Customers: $p < 0.05$). Although participants perceived the crimes involving \$5 million loss to be more harmful than the Low condition, these crimes were not perceived as more wrongful or serious with statistical significance (although the coefficients are in the expected direction).

H4b: People perceive cybercrimes as worse when large losses fall on consumers rather than on businesses.

Whether Acme or its customers bore the costs made little difference. Not only were the responses to the main Likert questions not statistically significant between the Acme High and Customer High conditions, the harmfulness of each condition was virtually the same ($\hat{\beta} = 0.03$, $se = 0.122$). This is somewhat surprising. I had expected that participants would empathize with customers over companies and that empathy would lead to ratings of damage to customers as more harmful than the same amount of damage to Acme. But this does not seem to have been the case. It could be that people are more sympathetic to customers than companies, as one might expect, but that the two conditions are not as similar as I had hoped. \$5 million in costs to a single company are not the same as \$5 million in costs spread among 100,000 people.

H5: An organization that had not patched its servers when it was breached is perceived as more co-responsible for the crime than an organization that had patched its servers.

The manipulation of whether Acme patched its servers had the expected effect on perceptions of the company’s partial responsibility for the crime. Participants found Acme more responsible for the crime when it had not patched its servers ($p < 0.01$). Participants did not find the crime significantly more wrongful, harmful, or serious in this case, suggesting that they distinguished between the seriousness of a crime and its causes.

Surprisingly, participants also rated the data as less sensitive when Acme had not patched its servers. Some people may have assumed that the data was poorly protected because it was less sensitive.

H6: Downloading data from a bank or government agency is perceived as more wrongful and harmful than downloading the same data from a non-profit.

The context manipulation showed no two-sided statistically significant effects on any of the main Likert questions except for how partially responsible the breached organization was. Participants judged the non-profit to be less responsible for the breach than they did the bank ($p < 0.01$) or the government agency ($p < 0.001$). Participants did rate the non-profit vignette as less serious than either the government or bank scenario with one-sided $p < 0.05$.

For the most part, Study I showed effects in the directions expected. Changing the data from directory information to health information increased perceived sensitivity. Increasing the number of records generally increased how wrongful, harmful, and serious the crime was seen. Interestingly, increasing the number of records also increased perceptions of how sensitive the data was. Cybercrime committed with a profit motive was rated as more wrongful than the same crime motivated by activism or a desire to learn. Respondents perceived an organization that had patched its servers to be less responsible for the crime than an organization that did not. The more costly a breach's consequences, the more harmful it was rated. Participants rated downloading data from banks and government agencies (and, in the factorial experiment, insurers) as more serious than downloading data from a non-profit.

Data sensitivity did not, however, appear to be a major component of seriousness. Despite the data sensitivity in Experiment 1 having the strongest effect of any manipulation, the perceived harmfulness, wrongfulness, and seriousness of the crime was not statistically significant across conditions.

The results of Study I support interpretations of seriousness as having components of both wrongfulness and harmfulness. Cybercrime vignettes that were rated as more wrongful were rated, with high significance, as more serious. So were vignettes that were rated as more harmful.

One of the more interesting results is the comparative reaction of the participants to cybercrimes committed by activists versus cybercrimes committed for profit. The former were considered significantly less blameworthy, and deserving significantly lighter sentences—contrary to the position sometimes taken by U.S. prosecutors.

Table 7 shows pairwise correlations between each dependent variable across all six between-subjects experiments in Study I.¹⁶² Wrongfulness, harmfulness, seriousness, and how harshly the crime should be punished are all positively correlated. The correlations between wrongfulness, harmfulness, and seriousness confirm previous work suggesting that the first two measures are components of the third.¹⁶³ The correlation between seriousness and how harshly the crime should be punished confirms that people want crimes that are more serious to be punished more harshly. Also unsurprising is the positive correlation between potential harm and measures of wrongfulness, harmfulness, seriousness, and punishment.

More interestingly, the results show a statistically significant positive correlation between perceived data sensitivity and ratings of the wrongfulness, harmfulness, seriousness, and harshness of punishment for crimes. This seems at odds with the results in the type-of-data experiment, which shows no significant effect on perceptions between medical data and directory data even though respondents rated the former as more sensitive than the latter. But the correlations are consistent with the results from the factorial experiment, as will be discussed in the next section.

¹⁶² Correlation matrices for each study do not differ meaningfully from the aggregate.

¹⁶³ See Warr, *supra* note 136, at 818–20.

Table 7: Pairwise correlation matrix for the DVs in the between-subjects experiments

	Wrongful	Harmful	Serious	Harshly	Sensitive	Responsible	Clever	Pot. Harmful
Wrongful	1.000							
Harmful	0.566***	1.000						
Serious	0.707***	0.614***	1.000					
Harshly	0.738***	0.669***	0.747***	1.000				
Sensitive	0.285***	0.296***	0.373***	0.298***	1.000			
Responsible	-0.027	0.031	0.039	-0.024	0.083***	1.000		
Clever	-0.051*	-0.044*	0.011	-0.067***	0.117***	0.113***	1.000	
Pot. Harmful	0.414***	0.399***	0.458***	0.413***	0.480***	0.077**	0.045	1.000

* p<0.05, ** p<0.01, *** p<0.001

Notes: The table shows pairwise correlations for the DVs across all six between-studies experiments. $N=2440$ for all pairings except those involving Pot. Harmful, for which $N=1618$.

As a final note on the correlation table, there are some statistically significant correlations involving how responsible ACR was for the crime and, separately, how clever the offender was. But the magnitudes of these correlations are tiny.

3.4. Study II: Factorial Vignette Survey Experiment

I followed the between-subjects experiments with an experiment using factorial vignette survey methodology. As discussed in Section 3.2.2, factorial vignette surveys are commonly used to study beliefs and normative judgments.¹⁶⁴ In this methodology, each participant rates a number of vignettes describing a scenario. The details of the scenario vary from vignette to vignette. In the parlance of factorial vignette methodology, the variables are known as dimensions and the possible values of those variables are called levels.

I decided to supplement the between-subjects experiments with a factorial vignette survey experiment for several different reasons. First, the factorial vignette methodology gives a better method of directly comparing the effects of different factors of a cybercrime. For example, we might want to know whether the scope or context of a cybercrime contributes more to perceptions of the seriousness of that crime. Because the between-subjects experiments were conducted at different times and, as between some experiments, with slightly different vignette texts, comparisons within a single experiment have more validity than those across the multiple experiments of Study I.¹⁶⁵

Second, because participants in a factorial vignette survey experiment each rate multiple vignettes, the factorial vignette methodology allows us to account for effects within subjects in addition to the between-subjects analysis. However, because the number of vignettes each participant rates must be kept reasonably small (twenty-five, in this case) to avoid fatigue, the statistical power of this analysis is limited.

¹⁶⁴ See KATRIN AUSPURG & THOMAS HINZ, FACTORIAL SURVEY EXPERIMENTS 13–15 (2015); Jasso, *supra* note 142, at 338–39; Rossi et al., *supra* note 77, at 62.

¹⁶⁵ See, e.g., Paul D. Allison, *Comparing Logit and Probit Coefficients Across Groups*, 28 SOC. METHODS & RES. 186 (1999); Carina Mood, *Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It*, 26 EUR. SOC. REV. 67 (2010).

Third, the different methodology lets us test the robustness of the results from the between-subjects experiments, obtain a larger sample size from a smaller number of participants (and thus gain greater statistical power without an accordant increase in cost), and refine some of the details of the rating task I asked participants to do.

Finally, the factorial vignette survey is a known methodology that has been used already in the literature on crime seriousness.¹⁶⁶

3.4.1. Methodology

3.4.1.1. Research Questions

The research questions are driven by the goals listed in Section 3.3.1.1. In terms of relative effect sizes, the results of the between-subjects surveys suggest that motivation—specifically, that of a profiteer versus a student or activist—is the largest factor in perceptions of cybercrime seriousness, followed by a crime’s consequences and scope. I conjectured that the same would be true when all were manipulated in the same study.

3.4.1.2. Design

The design for this study consisted of a factorial vignette survey experiment. I presented each participant with twenty-five vignettes describing a cybercrime scenario.¹⁶⁷ Each was structured as a paragraph describing the facts followed by a list of the factors that varied from one vignette to another.¹⁶⁸ The survey was similar in format to the between-subjects experiments, with some adjustments because participants would be asked to rate multiple vignettes.

The vignettes were of the following form:

Tom Smith is a computer programmer who looks for security flaws on the Internet. On September 3, 2014, Tom found a security flaw in the website of an organization named ACR and used that flaw to download records from ACR’s customer database. He anonymously released details about the flaw to the

¹⁶⁶ See generally, e.g., KATRIN AUSPURG & THOMAS HINZ, FACTORIAL SURVEY EXPERIMENTS 14 (2015); Larry A. Hembroff, *The Seriousness of Acts and Social Contexts: A Test of Black’s Theory of the Behavior of Law*, 93 AM. J. SOC. 322 (1987) (using the factorial methodology to study judgments of stabbing and theft scenarios); Jasso, *supra* note 142 (using the factorial survey methodology to study perceptions of five types of crimes); Rossi et al., *supra* note 77 (using the factorial survey methodology to study perceptions of fifty crimes).

¹⁶⁷ I would have preferred to present 40 vignettes per respondent, but a pilot study with that many vignettes showed signs of respondent fatigue, such as high dropout rates, and technical issues in the survey software. I therefore scaled back to 25 vignettes.

¹⁶⁸ Adopting a variation of Jasso’s terminology, I refer to the common story described in the vignettes as the “scenario,” a particular combination of that scenario with assigned values for each factor as a “vignette,” and the set of all vignettes that could be generated by the random selection of factor levels as the “vignette population.” See Jasso, *supra* note 142, at 340–41 (2006).

Internet, but did not use or release the records he downloaded. Before he did this, Tom had never been arrested or convicted of any crime.

ACR was *\$org*.

Tom downloaded *\$records* customer records.

Each record consisted of a customer's *\$data*.

Tom's motivation was to *\$motive*.

ACR spent *\$org_loss* to repair and secure its servers.

Its customers spent *\$cust_loss* each to protect themselves from identity fraud.

Tom was convicted of the crime and received a sentence of *\$sentence \$sent_type*.

I selected the values each variable could take to be the same as those used in the between-subjects experiments where possible. The values for each variable were:

- *\$org*: "a bank," "a non-profit organization," "an insurance company," "a government agency"
- *\$records*: 10, 100, 1,000, 10,000, or 100,000
- *\$data*: "e-mail address," "full name, phone number, and address," "full name, address, and social security number," "full name, health history, medical diagnoses, and prescription records," "full name, phone number, address, date of birth, and social security number," "full name, user ID, and password"
- *\$motive*: "learn about Internet security," "seek evidence of corporate corruption," "make money"
- *\$org_loss*: \$1000, \$10,000, \$100,000, \$1,000,000, \$10,000,000
- *\$cust_loss*: \$10, \$50, \$100, \$250, \$500
- *\$sentence*: 3 months, 6 months, 1 year, 2 years, 5 years
- *\$sent_type*: "probation," "in jail" (for sentences less than 1 year) or "in prison" (for sentences of one year or more)

The survey software selected the value of each variable randomly and independently for each vignette. Any given vignette therefore represented a random sample from the vignette population. The only exception to that independence is that I prevented health data ("full name, health history, medical diagnoses, and prescription records") from being selected as a data type when the organization type was a bank because participants might find it implausible that a bank would be holding health data in its database. I did not prevent other combinations that some participants might have found implausible, such as an organization suffering \$10 million in losses from the breach of 10 e-mail addresses (a combination that occurred 29 times in the data set). Treating the numerical factors *\$records*, *\$org_loss*, *\$cust_loss*, and *\$sentence* as continuous, the vignette population consisted of 138 vignettes. If the continuous variables were treated as categorical, the vignette population would contain 86,250 vignettes.

At the bottom of each vignette I presented a slider with the rating task asking participants to evaluate the sentence imposed. I limited the rating task to one question because of research showing that undesirable method effects increase when participants are asked multiple questions after each vignette.¹⁶⁹ The slider was anchored at each end with “Much too low” at the left and “Much too high” at the right. The marker on the slider was set to a starting position in the middle of the scale. The slider was unmarked except for the two anchors because of research suggesting that people tend to treat tick marks on a scale as “magnets”—a slider with five tick marks tends to be treated like a five-point Likert scale, for example.¹⁷⁰ Other research shows that adding numeric labels to a slider leads to increased rounding of responses.¹⁷¹ Figure 6 shows the slider scale.

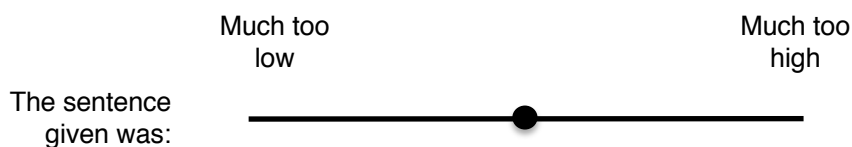


Figure 6: Factorial instrument rating task slider

I used a slider bar to approximate the real-number scale used in some previous factorial vignette surveys.¹⁷² The slider widget recorded a value from 0 to 256, with 0 corresponding to a rating that the sentence was “much too low” and 256 corresponding to a sentence that the participant believed was “much too high.” I normalized this to a 0 to 100 scale with 100 corresponding to a response that the punishment should have been higher—i.e., I reversed the scale as presented. I did not round to integer values when scaling.

After the instruction page, the survey presented participants with twenty-five vignettes, one per page, followed by the same attitude and demographic questions asked in the between-subjects surveys. Finally, the survey presented two open-ended questions: one asking participants what they thought the study was about and an optional question in which participants could enter comments about the study.

I ran mixed-effects regressions on the rating task, grouping by response ID. The regressions included controls for demographics, attention-check correctness, and privacy attitudes. As in the between-subjects studies, I treated gender, country of birth, age category, education, occupation, work situation, and the memory check variables as categorical variables. I treated as continuous variables the extent to which participants had been affected by cybercrime

¹⁶⁹ Katrin Auspurg & Annette Jäckle, *First Equals Most Important? Order Effects in Vignette-Based Measurement*, INST. SOC. & ECON. RESEARCH Working Paper 2012-01, (Jan. 18, 2012), <http://www.iser.essex.ac.uk/research/publications/working-papers/iser/2012-01>.

¹⁷⁰ See, e.g., Pete Cape, *Slider Scales in Online Surveys*, SURVEY SAMPLING INT’L (2009), http://www.websm.org/db/12/17947/Web_Survey_Bibliography/Slider_Scales_in_Online_Surveys/.

¹⁷¹ See, e.g., Mick P. Couper, Roger Tourangeau & Frederick G. Conrad, *Evaluating Effectiveness of Visual Analog Scales: A Web Experiment*, 24 SOC. SCI. COMPUTER REV. 227, 242 (2006).

¹⁷² See, e.g., Guillermina Jasso, *Exploring the Justice of Punishments: Framing, Expressiveness, and the Just Prison Sentence*, 11 SOC. JUST. RES. 397, 407–08; Rossi et al., *supra* note 77, at 66–67 (1985).

or privacy invasions and the extents to which they use fake personal information and are aware of media coverage of data misuse.

3.4.2. Theoretical Model

I use a multi-level model for respondents' belief function:

$$Y_{ij} = \beta_0 + \sum_k \beta_k X_{kij} + \sum_q \gamma_q Z_{qj} + u_j + \varepsilon_{ij} \quad (6)$$

where $i = 1 \dots n$ indexes the vignettes, $j = 1 \dots m$ indexes the respondents, $\beta_k X_{kij}$ are the vignette dimensions (scope, consequences, motivation, etc.) and coefficients, $\gamma_q Z_{qj}$ are respondent characteristics (gender, age, privacy attitudes, etc.) and coefficients, u_j is the respondent-specific error term, and ε_{ij} is the usual error term. This model allows for individual variation in intercepts and controls for respondent-level differences but assumes common slopes across respondents.¹⁷³ This assumption simplifies the model and lets us understand beliefs in the aggregate.

3.4.3. Results

I used MTurk to recruit participants 18 years of age or older who lived in the United States, had at least a 95% approval rating on MTurk, and had not previously participated in any of the studies described in this chapter. I screened potential participants to exclude anyone who had seen any of the between-subjects experiments or their pilots. Of 267 attempts to take the survey, there were 241 unique MTurk IDs (MIDs) and 224 completed responses. After removing one response because the participant answered that her age was under 18, a total of 223 responses remained (47% women; median age category 25–34).¹⁷⁴

¹⁷³ See Jasso, *supra* note 142, at 350–51.

¹⁷⁴ The 224 completed responses from 241 participants represent an abandonment rate of 7.1%. Two workers reported being unable to complete the survey because of technical issues. There was also a high retry rate; 17 completions were on a second attempt and 3 were on a third attempt. Fourteen people (5.8%) did not complete the survey and did not attempt to retake it. Three of them did not reach the first vignette, two stopped after two vignettes, and one stopped after four vignettes. Of the remaining eight participants who completed at least five vignettes but “abandoned” the survey, six completed at least fifteen questions and two completed all 25 questions and the CFIP questions but not the demographic questions. This pattern suggests that technical issues may have been responsible for many “abandoned” surveys even among MTurkers who did not try to retake the survey.

The distribution of responses shows signs of censoring and clustering at the midpoint. About 10% of all ratings were at the midpoint of the slider. Another 5% were at the left end (“Much too low”) and 3% were at the high end (“much too high”). Respondents who answered the attention-check question correctly gravitated to the midpoint and extremes slightly less often than those who did not, 17% to 22% (a statistically significant difference at $p < 0.001$, χ^2). Censored and clustered responses were not distributed equally among participants. About 11% of respondents (25) rated 10 or more of the 25 vignettes at the extremes or middle, and 7% (13) rated at over half of their vignettes that way. One person rated all vignettes either at the bottom (20 times) or middle (5 times).

Table 8: Mixed-effects regression for the factorial experiment

	Betas	se
log(Records)	0.584***	(0.073)
log(Org Loss)	0.640***	(0.096)
log(Cust Loss)	0.672***	(0.184)
Organization (vs. Bank)		
Government	-1.187	(0.755)
Non-profit	-1.563*	(0.795)
Insurer	-0.846	(0.781)
Data (vs. E-mail)		
Name, addr, SSN	10.112***	(1.110)
Name, health history, diagnoses, prescriptions	11.104***	(1.149)
Name, phone, addr, DOB, SSN	11.723***	(1.176)
Name, phone, addr	6.213***	(0.907)
Name, user ID, pwd	6.682***	(1.012)
Motivation (vs. Profiteer)		
Student	-10.445***	(0.941)
Activist	-10.573***	(0.958)
log(Sentence)	-8.729***	(0.458)
Probation	7.519***	(1.424)
log(Sentence) \$\$ Probation	2.449***	(0.545)
Female	2.642	(1.546)
US birth	2.188	(2.757)
CFIP score	0.980	(1.007)
Freq. aff by cybercrime	0.367	(1.304)
Media awareness	-0.342	(0.515)
Attn. check	1.307	(1.809)
_cons	49.567***	(7.448)
sd(_cons)	10.264***	(0.703)
sd(Residual)	17.844***	(0.392)
<i>N</i>	5575	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The table shows mixed model regression results for responses to the factorial experiment. The DV for each regression is the rating of punishment severity normalized to a 100-point scale. Higher numbers correspond to beliefs that punishments should be harsher.

Table 8 shows the results of the mixed effects regressions on the 100-point normalized rating task. The results are robust to exclusion of answers from participants who did not answer the attention check question correctly.¹⁷⁵

All of the factors show statistically significant effects for at least some values. The strongest effect in terms of magnitude is the difference between the student (or activist) and profiteer motivations. A vignette in which the offender's motive was profit received a rating that was a little more than 10 points higher on the 100-point scale than the motive for a student or activist.

¹⁷⁵ In a regression without incorrect attention check answers, the coefficient for log(\$cust_loss) drops in significance ($\beta = 0.51$, $p < 0.05$, $se = 0.21$) and the coefficient for non-profit as the organization type drops out of significance ($\beta = -1.43$, $se = 0.84$). All other coefficients retain their significance (or lack thereof) and have similar values.

The next highest effect is the type of data. This is somewhat surprising because the type of data was not a statistically significant manipulation in the between-subjects studies. Of course, much of that is because the low-sensitivity data type in the factorial study consists only of e-mail addresses instead of names, phone numbers, and addresses as in the between-subjects experiment. But even between the two data types used in the between-subjects experiment (name, phone number, and address versus health data), there is a statistically significant effect of about $\beta = 4.9$ ($p < 0.001$, $se = 0.97$) in the factorial study. Some of the difference in results might be explained by the larger sample size, but the effect sizes in the between-subjects experiments were very small—the coefficients for seriousness and harshness of punishment in the type-of-data experiment were roughly an order of magnitude lower than those in the motivation experiment, for example. Thus it does not seem likely that effect size alone accounts for the difference.

As in the between-subjects experiments, the scope (number of records) and consequences (loss to customers and the breached organization) are significant but with small effect magnitudes. Note, however, that because the explanatory variables are log transformed, the effect sizes are not quite as tiny as they appear at first glance in the regression table. Increasing the loss to the breached organization or customers by a factor of ten would correspond to an increase of about 1.5 points in the scaled rating. A tenfold increase in the number of records would correspond to an increase of 1.3 rating points on the 100-point scale according to the results in the main model; an increase along the full 10–1,000,000 record range would be expected to add about 6.7 points. This is still a relatively small effect: all else being equal, the increase in perceived seriousness from a breach of 1,000,000 records instead of 10 records is about the same as the difference between a breach of names, user IDs, and passwords instead of e-mail addresses.

I found no statistically significant interaction effects. I also checked interactions for other combinations of explanatory variables and found no statistically significant interactions.

3.5. Discussion

3.5.1. Comparison of Results Between the Two Studies

Although the results from the two studies were mostly similar, some interesting differences do appear.

Data sensitivity did not appear to be a major component of perceived cybercrime seriousness in Study I. But the factorial experiment showed some significant effects between broad categories of data types. Crimes in which only e-mail addresses were accessed were rated as deserving of significantly less harsh punishments. The other five data types in the factorial experiment showed something of a partitioning. Data involving either health data or Social Security Numbers had the largest coefficients. The middle tier includes (1) directory information and (2) usernames and passwords, which have roughly the same coefficients. This is surprising,

because phone numbers would seem to be less potentially harmful than usernames and passwords.

But this result may simply be an effect of the *length* of the data type description. Running the basic model (1) regression from Table 8 with the length of the data type string (as a continuous variable) instead of the data type categorical variable results in a coefficient for the string length ($\beta = .165$, $se = .015$) that is also statistically significant at $p < 0.001$. Multiplying this coefficient by the number of characters in each data type results in numbers that are, with the exception of “Name, address, and SSN,” not far from those in model (1) in Table 8.¹⁷⁶ Perhaps respondents used the length of the data type as a heuristic. Unfortunately, because the length of the data type descriptions and the sensitivity of the data listed are not independent, it is impossible to disentangle their effects in the results.

3.5.2. Implications for Sentencing Policy

The factorial vignette survey experiment showed a marked disparity between the effect of a breached organization’s loss on perceptions of crime severity and the impact of loss on sentences. The main factorial regression equation predicts that increasing the organization’s loss from \$1,000 to \$10,000,000 corresponds to a 5.9-point increase in severity rating (on a 100-point scale). The same change in dollar amount would lead to a 20-point increase in offense level in the 2018 U.S. Sentencing Guidelines,¹⁷⁷ enough to bump the presumptive sentencing range for a first offense with no other enhancements from 0–6 months to 63–78 months.¹⁷⁸ For comparison, the coefficient on $\$sentence$ (when $\$probation = 0$) is -8.729, which means the modeled decrease in 100-point rating from a 3 month to 5 year sentence is -26.15. In other words, the actual increase in presumptive punishment from the increased amount of loss is about three times what respondents in the experiment rate as appropriate.

Motivation was much more important in the results than it is in sentencing. Respondents judged crimes with a profit motive to be much more serious than those committed for activism or curiosity. The coefficient of roughly -10.5 in the main regression for the Student and Activist levels of motive means that the Profiteer motive increases the rating of a cybercrime by about the same amount as more than tripling a prison or jail sentence (a factor of 3.3, to be more precise). That suggests that there could be support for increasing a 3-month sentence to 10 months or a 12-month sentence to 40 months when profit is the motive for the crime (or, alternately, that crimes committed for motives other than profit should be discounted by reversing those numbers). That increase in sentence duration would correspond to an increase of about 8 to 10 offense levels in the sentencing guidelines.

¹⁷⁶ Reading down the column: 7.8, 11.6, 12.4, 5.9, and 5.3.

¹⁷⁷ See U.S. SENTENCING GUIDELINES MANUAL §2B1.1(b)(1).

¹⁷⁸ *Id.* at §5.A. Note, however, that some sections of the CFAA carry maximum sentences of 5 years for a first offense.

The type of organization was not a statistically significant factor in evaluations of crime seriousness. This stands in contrast to the CFAA’s specific provisions covering financial and government information,¹⁷⁹ or government computers.¹⁸⁰

Table 9 lists the effect of offense factors on perceptions and sentencing. For example, the model predicts that a cybercrime with a loss of \$10,000,000 instead of \$1000 would increase perceptions of the seriousness of that crime by 8.8 points on the 100-point scale (all other factors held fixed at the mean). The recommended sentence, however, would be 91 to 113 months longer (though maximum sentences might reduce that).

Table 9: Impact of offense factors on perceptions and sentences

Factor	Range	Empirical effect	Sentencing effect
Records (Scope)	100,000 vs. 10	+8.7	Depends on amount of cust. loss
Org. Loss	\$10,000,000 vs. \$1000	+8.8	+91–113 months
Cust. Loss (each)	\$500 vs. \$10	+3.5	Depends on no. of records
Motivation	Profiteer vs. Activist	+10.6	5 year max sentence
Context	Bank vs. Non-profit	+1.6	5 year max sentence
Type of Data	Name, phone, addr, DOB, SSN vs. e-mail	+11.7	+4–6 months

Notes: Empirical effect is based on coefficient estimates in the factorial experiment, assuming all other factors held fixed. Sentencing effect assumes criminal history category of I, 6 point base offense level, and two 2-point enhancements for sophisticated means and use of a special skill, for an offense level of 10 and sentencing range of 6–12 months.

To illustrate in more concrete terms the differences between perceptions of cybercrime seriousness and how the sentencing guidelines weigh the attributes of a cybercrime, consider the hypothetical crime I used in my experiments: a person named Tom Smith discovers a security flaw in a website and uses that flaw to access a company’s internal network and download records containing personal information. The experimental results show that people perceive a computer crime to be more serious when the data is more sensitive, the offender is motivated by financial gain, the amount of loss is high, and a large number of records are affected—in roughly that order. If sentencing reflected public perceptions, a crime with these features would be punished more harshly than a crime in which these factors are less true.

Suppose the hypothetical Tom’s motivation was to make money, that the number of records was 100,000, and that the data contained full names, addresses, phone numbers, dates of birth, and social security numbers. All these parameters are the highest values for factors deemed important in the experiments. Assume losses by customers were minimal (because Tom did not release the data) or cannot be proven and that ACR was a non-profit. The maximum sentence would be five years because the offense was committed for purposes of financial

¹⁷⁹ 18 U.S.C. § 1030(a)(2)(A)–(B).

¹⁸⁰ *Id.* § 1030(a)(3).

gain.¹⁸¹ Also, the value of the records Tom obtained may well be worth more than \$5000.¹⁸² The base offense level under section 2B1.1 would be 6.¹⁸³ The enhancements for using special skill¹⁸⁴ or sophisticated means,¹⁸⁵ which seem to be common in CFAA cases, add two points each. Because the data Tom obtained included personal information, another two-point enhancement applies.¹⁸⁶ If ACR's only loss is spending \$1000 to repair and secure its servers, no enhancement for the amount of loss applies and the total offense level (assuming no other adjustments apply) is 12—which corresponds to a presumptive sentencing range of 10 to 16 months at criminal history category I.

Now assume a different set of facts from the experiments. In this version, Tom was an activist (perceived as less serious than the profiteer, all other factors held constant, by 10.5 points on the 100-point scale), he downloaded 1,000 records (2.7 points less serious), and the data contained only e-mail addresses (11.7 points less serious than the information in the facts above). The maximum sentence is likely one year instead of five: the offense was not committed for financial gain and the value of 1,000 e-mail addresses is far less than \$5,000,¹⁸⁷ so the higher maximum sentence applies only if “the offense was committed in furtherance of any criminal or tortious act in violation of the Constitution or laws of the United States or of any State.”¹⁸⁸ If ACR spent \$1,000 as in the previous fact pattern, the offense level would be 10 (assuming e-mail addresses alone are not “personal information” as defined in the guidelines),¹⁸⁹ which corresponds to a sentence of 6 to 12 months—a reduction of 2 offense level points and four months of presumptive sentence.

Next, consider the possible sentences if ACR responded to Tom's hack by hiring consultants and investigators and notifying all 1,000 customers of the breach by regular mail and phone calls, at a cost of \$300,000. The perceived severity of the crime would increase due to the

¹⁸¹ See *id.* § 1030(c)(2)(B)(i).

¹⁸² One study found that a full set of personal information including SSN, address, and birthdate had a median price of \$21 on the “dark web.” See Keith Collins, *Here's What Your Stolen Identity Goes For on the Internet's Black Market*, QUARTZ (July 23, 2015), <https://qz.com/460482/heres-what-your-stolen-identity-goes-for-on-the-internets-black-market/>. Others found that bulk data sells for pennies per record. See Itay Glick, *Darknet: Where Your Stolen Identity Goes to Live*, DARK READING (Aug. 19, 2016), <http://www.darkreading.com/endpoint/darknet-where-your-stolen-identity-goes-to-live/a/d-id/1326679>; Brian Krebs, *How Much is Your Identity Worth?*, KREBS ON SECURITY (Nov. 8, 2011), <https://krebsonsecurity.com/2011/11/how-much-is-your-identity-worth/>. Even at a nickel per record, however, a set of 100,000 records would be worth \$5000.

¹⁸³ U.S. SENTENCING GUIDELINES MANUAL § 2B1.1(a).

¹⁸⁴ *Id.* § 3B1.3.

¹⁸⁵ *Id.* § 2B1.1(b)(10)(C).

¹⁸⁶ *Id.* § 2B1.1(b)(17).

¹⁸⁷ In 2011, one could buy a million e-mail addresses for \$25. Carlton Purvis, *\$00.000025: The Going Rate on the Black Market for Your Email Address*, SECURITY MGMT. (Aug. 26, 2011), <https://sm.asisonline.org/Pages/00000025-going-rate-black-market-your-email-address-008950.aspx>.

¹⁸⁸ 18 U.S.C. § 1030(c)(2)(B)(ii).

¹⁸⁹ The sentencing guidelines define “personal information” as “sensitive or private information involving an identifiable individual (including such information in the possession of a third party), including (A) medical records; (B) wills; (C) diaries; (D) private correspondence, including e-mail; (E) financial records; (F) photographs of a sensitive or private nature; or (G) similar information.” U.S. SENTENCING GUIDELINES MANUAL § 2B1.1 cmt.1.

larger loss by a mere 3.65 points on the 100-point scale, but the offense level would more than double, to a total of 22.¹⁹⁰ The presumptive range would be 41–51 months with a statutory maximum of one year. The weight the guidelines place on loss under section 2B1.1 greatly outdistances not only the increase in perceived severity resulting from the greater loss but also the statutory maximum. And two facts that contributed little or nothing to the offense level in the previous fact pattern—the motive and value of the information obtained—turn out to be critical threshold issues. Changing the motive from activism to financial gain or the value of the data from sub-\$5,000 to more than \$5,000 can change a one-year maximum sentence to a recommended sentence of at least three and a half years.

Finally, assume the first set of facts again: profit motive, 100,000 records, and data consisting of full names, addresses, phone numbers, dates of birth, and social security numbers. But as in the previous example, ACR spent \$300,000 reacting to the incident. The offense level would be 24: 12 as in the first fact pattern plus 12 for the amount of loss. The recommended sentencing range is 51 to 63 months. Because the motive is financial gain and the records consist of personal information, the maximum sentence is five years.

Table 10: Sentencing examples for the factorial scenario

	Loss: \$1000	Loss: \$300,000 (+3.65)
Motive: Profiteer (+10.5)	Offense level: 12	Offense level: 24
Scope: 100,000 records (+2.7)	Guideline range: 10–16	Guideline range: 51–63
Records: Name, addr, phone no., DOB, SSN (+11.7)	mo. Max: 5 years	mo. Max: 5 years
Motive: Activist	Offense level: 10	Offense level: 22
Scope: 1,000 records	Guideline range: 6–12 mo.	Guideline range: 41–51
Records: E-mail addresses	Max: 1 year	mo. Max: 1 year

Note: The table lists offense levels, recommended sentencing ranges, and maximum sentences for the fact values listed. Values in parentheses are the modeled change, on a 100-point scale, in perceived severity compared to the lower level, assuming all other factors are held fixed at the mean (e.g., a loss of \$300,000 is modeled as 3.65 points higher on the 100-point scale than a loss of \$1000).

Two lessons can be gleaned from these examples (which Table 10 summarizes). First, as mentioned, the amount of loss has an outsized effect on recommended sentences compared to the importance of that factor on perceptions of crime seriousness. A change in loss that increases the perceived seriousness of a crime by less than 4 points on a 100-point scale can increase the recommended sentencing range from 10–16 months to 51–63 months. Second, because motive and the sensitivity of the data can increase maximum sentences but have only minimal effect on calculations under the guidelines, their impact primarily depends on whether a prosecutor can find other ways (such as charging additional crimes to create “another offense” or by coming up with creative valuations of data) to increase the maximum sentence.

¹⁹⁰ U.S. SENTENCING GUIDELINES MANUAL § 2B1.1(b)(1) (listing a 12-point increase in offense level for an offense with more than \$250,000 in loss).

Apart from the language about gaining access to the company's internal network, the hypothetical is similar to the facts of the case against Andrew "Weev" Auernheimer, who discovered a vulnerability in AT&T's web site for iPad registrations and downloaded more than 100,000 records.¹⁹¹ Auernheimer was convicted of conspiracy and identity fraud.¹⁹² He received a sentence of 41 months that was overturned on jurisdictional grounds.¹⁹³

The government argued for an offense level of 20, which carried a presumptive sentencing range of 33–41 months. The offense level was based on a base offense level of 6; three 2-point enhancements for use of a special skill, use of sophisticated means, and dissemination of personal information; and an 8-point enhancement for a loss of \$73,000 incurred by AT&T in mailing notices to affected customers.¹⁹⁴ The base offense level and enhancements for special skills and sophisticated means accounted for ten offense levels, corresponding to a presumptive sentencing range of 6–12 months. The two-point enhancement for use of a special skill alone would have increased that to 10–16 months. The enhancement for amount of loss would have increased the guidelines range from 6–12 months to 27–33 months. Thus, the amount of loss—the \$73,000 AT&T spent notifying customers—increased Auernheimer's presumptive sentence five times more than the type of data did.¹⁹⁵ Note, however, that the fact that Auernheimer was accused of accessing identifying information with the intent to commit a violation of federal law allowed him to be prosecuted under the identity theft statute, which is also sentenced under section 2B1.1 of the guidelines but carries a five-year maximum sentence. Had he been charged under the CFAA, the government would have had to show that the value of the information Auernheimer obtained was more than \$5,000.¹⁹⁶

As mentioned in Section 3.2.1.2, most CFAA offenses are sentenced under section 2B1.1 of the guidelines, which covers economic crimes such as fraud and larceny. The results support existing arguments that this is a poor fit.¹⁹⁷ The heavy reliance that section 2B1.1 places on the amount of loss in calculating a recommended sentence is not reflected in public perceptions. Meanwhile, factors that the respondents do rate as important, such as motive, type of data, and scope, are barely factors in 2B1.1.

¹⁹¹ See Kim Zetter, *AT&T Hacker 'Weev' Sentenced to 3.5 Years in Prison*, WIRED (Mar. 18, 2013), <https://www.wired.com/2013/03/att-hacker-gets-3-years/>.

¹⁹² Judgment in a Criminal Case, *United States v. Auernheimer*, No. 2:11-cr-470 (D.N.J. Mar. 19, 2013).

¹⁹³ *Id.*

¹⁹⁴ Letter from United States, *Auernheimer*, No. 2:11-cr-470 (D.N.J. Mar. 15, 2013).

¹⁹⁵ As Orin Kerr notes, "the Guidelines recommended two extra years in jail because AT&T opted to mail out a postal letter." Kerr, *supra* note 113, at 1557–58.

¹⁹⁶ See 18 U.S.C. § 1030(c)(2)(b)(iii).

¹⁹⁷ See Kerr, *supra* note 113, at 1554–56.

3.5.3. Limitations and Opportunities for Further Research

I emphasize, as Rossi, Simpson, and Miller did in 1985,¹⁹⁸ that I do not claim sentences should be determined by public opinion. As mentioned at the beginning of this chapter, lay opinions of sentencing are subject to biases, lack of information, and misperceptions. But these perceptions do inform public policy decisions. When perceptions are wildly out of line with sentencing mechanisms, it is at least worth asking whether those mechanisms are truly aligned with public policy objectives (and if not, why). Furthermore, my measurement of perceptions is focused on the relative importance of various factors rather than on the comparison of total sentences.

The experiments I have discussed are all based on vignettes describing a data breach. But there are many types of cybercrime, including payment card fraud, scamming, online banking fraud, phishing, and viruses. A natural extension of this work would be to compare different types of cybercrime.

Another limitation of this work is that it ignores many victim and offender characteristics, other than the offender's cleverness. The victims in the scenarios are limited to a corporation and generic data subjects. But victim characteristics may be important too. Although other offender and victim characteristics should not bias these results, assuming these unobserved characteristics and participant assumptions about them were distributed randomly, it is possible that the effects I do measure are smaller than those I chose to ignore.

Because I use MTurk for the respondent sample, the results should not be considered representative of the U.S. population at large. Although MTurk studies have been shown to be better than most "samples of convenience," biases may exist within the MTurker community that affect the results.

The surprising appearance of data sensitivity among statistically significant results of other manipulations suggests that perceptions of data sensitivity might be another area for future research. The public's perceptions of fault on the part of breached organizations is another area of possible further study.

Finally, although the studies described in this chapter support the argument that most computer crimes should not be sentenced as fraud crimes, the results say nothing about whether trespass is the correct analogue. Computer crimes also have features of burglary, for example. The next chapter explores this further.

3.5.4. Conclusion

An attacker's motivation, the type of data affected, and the amount of loss are all statistically significant factors in perceptions of the seriousness of a Computer Fraud and Abuse Act crime. Sentencing under the Act places tremendous weight on the amount of loss. But that weight is not reflected in public attitudes. Another factor in sentencing—the target of the crime—appears to have no statistically significant effect on perceptions. In contrast, the most

¹⁹⁸ Rossi et al., *supra* note 77, at 61.

important factor in public ratings of crime seriousness is the attacker's motivation, which has a much less drastic impact in the sentencing guidelines.

I stress again that sentences should not be determined solely by public opinion. But if the criminal codes "reflect through the state legislature's deliberations and actions some understanding, however dim and remote, of what 'the public' deems appropriate for the crimes in question,"¹⁹⁹ it is reasonable to ask whether those reflections are distorted. This research suggests that they are.

¹⁹⁹ *Id.* at 59–60.

4. An Empirical Analysis of Sentencing and Perceptions of “Access to Information” Computer Crimes

4.1. Introduction

The previous chapter shows that the factors that contribute to CFAA sentencing are misaligned with perceptions of how those factors contribute to the seriousness of the crimes. This misalignment implies that the CFAA may be misplaced as a fraud crime sentenced under section 2B1.1 of the sentencing guidelines. But if CFAA crimes—in particular, access-to-information crimes under subsection (b)(2)—should not be sentenced according to section 2B1.1, where do they belong? Would (a)(2) fit under the guidelines for other crimes? Or would a brand new set of guidelines specific to computer crimes make more sense?

The question of how computer crimes should be sentenced also raises a fundamental issue that has long been debated but which has not been explored empirically: just what *is* computer crime, anyway? Is it basically the online version of trespass, as has frequently been argued (or assumed) in case law, scholarly commentary, and the CFAA’s legislative history?²⁰⁰ Or does the value of the information accessed in an (a)(2) crime make it more like the crime of burglary? Is fraud actually a closer analogue in terms of the perceptions of seriousness? Or is the CFAA different enough from all of these that trying to fit it into the sentencing regime for any physical-world crime is doomed to the same sort of problems it has now?

The answers to these questions are not merely philosophical. A disconnect between the “true nature” of computer crime and its sentencing has real-world implications in how computer crimes are punished. It also affects whether computer crime punishment is perceived as being fair—and indeed, there seems to be a widespread perception that computer crime sentencing is often harsher than it ought to be.²⁰¹

²⁰⁰ See *infra* Section 4.2.1.

²⁰¹ See, e.g., Hanni Fakhoury, *The Matthew Keys Case, the CFAA, and Why Maximum Sentences Matter*, EFF (Mar. 14, 2013), <https://www.eff.org/deeplinks/2013/03/3-months-or-35-years-understanding-cfaa-sentencing-part-1-why-maximums-matter> (“[T]his case underscores how computer crimes are prosecuted much more harshly than analogous crimes in the physical world”); Molly Sauter, *Online Activism and Why the Computer Fraud and Abuse Act Must Die*, BOINGBOING (Sept. 26, 2014), <https://boingboing.net/2014/09/26/fuckthecfaa.html> (“Potential sentences for DDoS actions in the United States are high compared to other crimes and especially compared to other types of traditionally recognized activist activities”); Gautham Nagesh, *Congress: Is Computer Law Too Tough on Hackers?*, ROLL CALL (Feb. 13, 2013), <https://www.rollcall.com/2013/02/13/congress-is-computer-law-too-tough-on-hackers/> (quoting EFF legal director Cindy Cohn as saying, “CFAA penalties are out of proportion with the actual offenses”); James Hendler, *It’s Time to Reform the Computer Fraud and Abuse Act*, SCIENTIFIC AMERICAN (Aug. 16, 2013), <https://www.scientificamerican.com/article/its-times-reform-computer-fraud-abuse-act/> (comparing CFAA sentences to those for child sex abuse, gang-related homicide, and child pornography); Tim Wu, *Fixing the Worst Law in Technology*, NEW YORKER (Mar. 18, 2013), <https://www.newyorker.com/news/news-desk/fixing-the-worst-law-in-technology> (“[The CFAA] can now put people in prison for decades for actions that cause no real economic or physical harm”); Decian McCullagh, *From ‘WarGames’ to Aaron Swartz: How U.S. Anti-Hacking Law Went Astray*, C|NET (Mar. 13, 2013), <https://www.cnet.com/news/from-wargames-to-aaron-swartz-how-u-s-anti-hacking-law-went-astray/> (“[The CFAA] has become the proverbial hammer where a scalpel will do”).

This chapter seeks to shed an empirical light on the question of whether (a)(2) is punished and perceived like a trespass, burglary, or fraud statute. I do this using two studies: (1) an analysis of real-world sentencing data, and (2) an MTurk experiment to measure perceptions.

To analyze sentencing, I built a custom data set by collecting and coding fact patterns from court filings and combining that information with data from the U.S. Sentencing Commission's files for individual offenders. The resulting data set improves on previous analyses of the CFAA by its completeness: the data contains facts, offender characteristics, and sentencing calculations for 1,095 CFAA sentences (including 572 sentences under (a)(2)), which represents 96% of all CFAA sentences imposed from 2005 through 2018.

I analyzed perceptions using an experimental survey of 499 participants on Amazon Mechanical Turk. The survey asked participants to rate the harmfulness, wrongfulness, and seriousness of 28 short vignettes describing federal computer, trespass, burglary, and fraud crimes. The vignettes were drawn from the elements of the crimes as defined in their statutes and fact patterns found in actual cases. I set the monetary loss (if any) for each crime so that the resulting offense levels under the sentencing guidelines could be used as controls.

The results suggest that (a)(2) is not like trespass, burglary, or fraud, at least in terms of current punishments or perceptions. CFAA (a)(2) crimes receive lower punishments and are perceived to be less serious crimes than fraud or burglary crimes. But (a)(2) crimes receive harsher punishments and are perceived to be more serious than federal trespass crimes that do not involve weapons. These results may lend support to arguments that CFAA sentencing should be covered under its own section of the sentencing guidelines, at least for (a)(2) offenses.

4.2. Background and Related Work

4.2.1. Computer Crime Norms

Of the seven overlapping forms of conduct that are prohibited under the CFAA, subsection (a)(2) is the most frequently charged and the broadest in terms of the type of conduct it covers.²⁰² The text of this subsection provides for criminal penalties against anyone who “intentionally accesses a computer without authorization or exceeds authorized access, and thereby obtains information” from virtually any computer.²⁰³

Subsection (a)(2) suffers from something of an identity crisis. Its purpose, according to the Senate Committee report amending the CFAA in 1986, was “the protection, for privacy

²⁰² See *infra* Table 11.

²⁰³ 18 U.S.C. § 1030(a)(2). Section (a)(2) applies to information obtained from any “protected computer,” which includes “at a minimum . . . all computers that connect to the internet.” *Van Buren v. U.S.*, 593 U.S. ___, slip op. at 2 (2011). In its original form, section (a)(2) only prohibited access to information in financial records. Amendments in 1996 added the language in (a)(2)(B) and (a)(2)(C), extending the coverage of subsection (a)(2) to virtually all computers. Because a “protected computer” includes any computer “which is used in or affecting interstate or foreign commerce or communication,” 18 U.S.C. § 1030(e)(2)(B), and because the unauthorized access need not take place across state lines—it is enough that the computer have access to the Internet—nearly all computers meet the definition of “protected computers.”

reasons, of computerized credit records and computerized information relating to customers' relationships with financial institutions."²⁰⁴ Its language, which prohibits unauthorized access to information, makes it look a little like a privacy law.²⁰⁵ But the sentencing guidelines treat it as a fraud crime—the statutory index to the guidelines calls for 18 U.S.C. § 1030(a)(2) to be sentenced according to § 2B1.1,²⁰⁶ covering various forms of fraud and theft.²⁰⁷ And analogies have been drawn, particularly among legal scholars and in the legislative history of the CFAA, to trespass, breaking and entering, and burglary.²⁰⁸

The argument for section (a)(2) being a fraud crime is largely a matter of lexicography, proximity, and history: it is part of a law titled the “Counterfeit Access Device and Computer Fraud and Abuse Act of 1984,”²⁰⁹ it was codified in the section of the criminal code devoted to “Fraud and False Statements,”²¹⁰ and it appears next to other subsections of the CFAA that more clearly resemble fraud. Section (a)(4) explicitly applies to unauthorized access with the intent to defraud,²¹¹ and section (a)(5), covering damage to computers, can also plausibly be treated as an economic crime, at least if the amount of loss can reasonably be quantified.²¹² And because the language of (a)(2) was originally much narrower, applying only to unauthorized access (or access for an unauthorized purpose) to financial or consumer reporting records, even (a)(2) originally looked more like a fraud statute than it does now.²¹³ Thus, as Orin Kerr explains, when the Sentencing Commission decided which section of the guidelines should apply to the

²⁰⁴ S. Rep. No. 99-432, at 5 (1986), *reprinted in* 1986 U.S.C.C.A.N. 2479, 2484.

²⁰⁵ There is some evidence in the legislative history for this interpretation. *See* S. Rep. No. 99-432, at 6 (1986), *reprinted in* 1986 U.S.C.C.A.N. at 2484 (“The premise of [subsection (a)(2)] is privacy protection . . .”).

²⁰⁶ U.S. SENTENCING GUIDELINES MANUAL, appx. A, at 553 (U.S. SENTENCING COMM’N 2018).

²⁰⁷ The full heading of § 2B1.1 is “Larceny, Embezzlement, and Other Forms of Theft; Offenses Involving Stolen Property; Property Damage or Destruction; Fraud and Deceit; Forgery; Offenses Involving Altered or Counterfeit Instruments Other than Counterfeit Bearer Obligation of the United States.”

²⁰⁸ *See, e.g.,* Orin S. Kerr, *Norms of Computer Trespass*, 116 COLUMBIA L. REV. 1143 (2016) (applying trespass norms to unauthorized access to computers); Orin S. Kerr, *Trespass, Not Fraud: The Need for New Sentencing Guidelines in CFAA Cases*, 84 GEO. WASH. L. REV. 1544 (2016); Lenese C. Herbert, *Cybercrimes and Hacking Issues*, ABA-ALI COURSE OF STUDY: INTERNET LAW FOR THE PRACTICAL LAWYER, SK102 ALI-ABA 139 (2005) (comparing cybercrime to “a number of traditional crimes” including “trespass, burglary, breaking and entering, mail fraud, extortion, [and] embezzlement”). *See also* Richard A. Epstein, *Cybertrespass*, 70 U. CHI. L. REV. 73 (2003) (comparing computer intrusions without damage to civil trespass outside the CFAA context).

²⁰⁹ Pub. L. No. 98-473, ch. XXI, 98 Stat. 1837, 2190 (1984)

²¹⁰ 18 U.S.C. ch. 47.

²¹¹ *See* 18 U.S.C. § 1030(a)(4) (establishing penalties for anyone who “knowingly and with intent to defraud, accesses a protected computer without authorization, or exceeds authorized access, and by means of such conduct furthers the intended fraud and obtains anything of value, unless the object of the fraud and the thing obtained consists only of the use of the computer and the value of such use is not more than \$5,000 in any 1-year period”).

²¹² Of course, one of the major complaints about the CFAA is that the amount of loss is not reasonably quantified. *See supra* note 71.

²¹³ *See* Pub. L. No. 98-473, § 2102, sec. 1030(a)(2), 98 Stat. 1837, 2190–91 (1984) (establishing penalties for anyone who “knowingly accesses a computer without authorization, or having accessed a computer with authorization, uses the opportunity such access provides for purposes to which such authorization does not extend, and thereby obtains information contained in a financial record of a financial institution . . . or contained in a file of a consumer reporting agency on a consumer . . .”).

CFAA, it probably thought that § 2B1.1 was a reasonable fit.²¹⁴ But the scope of (a)(2) has since been broadened.

The CFAA has often been compared to a trespass crime. Orin Kerr, for example, has argued that “[m]ost CFAA offenses are trespass offenses, not economic crimes.”²¹⁵ Josh Goldfoot and Aditya Bamzai applied a trespass framework for understanding when access should be considered to be unauthorized, arguing that the standard for “authorization” under the CFAA should be the same as that for physical trespass.²¹⁶ Susan Brenner, arguing that there is no such thing as a “cybercrime,” stated that criminal trespass is an “obvious” analogy to hacking.²¹⁷

Courts have also described the CFAA as a trespass crime. Sometimes this has been through explicit analysis.²¹⁸ The most prominent recent example is Justice Thomas’s dissent in *Van Buren v. United States*, which analyzed authorization in terms of “basic principles of property law” by analogy to trespass, theft, and bailment.²¹⁹ At other times, CFAA-as-trespass is mentioned in passing, as if the comparison is obvious.²²⁰ But not all courts have accepted the trespass analogy.²²¹

The legislative history of the CFAA and its amendments also refers at times to some of the conduct the statute criminalizes as “trespass.”²²² The Senate report to the 1986 Computer Fraud and Abuse Act, for example, described section 1030(a)(3) as “a simple trespass offense.”²²³ The report also distinguished between the “acts of fraud under (a)(4), punishable as felonies, and acts of simple trespass, punishable in the first instance as misdemeanors.”²²⁴ The House report of the same year described hackers as “trespassers, just as much as if they broke a window and crawled into a home while the occupants were away.”²²⁵

²¹⁴ See Kerr, *Trespass, Not Fraud*, *supra* note 208, at 1547.

²¹⁵ *Id.* at 1545. See also Kerr, *Norms of Computer Trespass*, *supra* note 208.

²¹⁶ Josh Goldfoot & Aditya Bamzai, *A Trespass Framework for the Crime of Hacking*, 64 Geo. Wash. L. Rev. 1477 (2016).

²¹⁷ Susan W. Brenner, *Is There Such a Thing as a “Virtual Crime?”*, 4 CAL. CRIM. L. REV. 1, 80 (2001).

²¹⁸ See, e.g., *United States v. Valle*, 807 F.3d 508, 525 (2d Cir. 2015) (interpreting “authorization” in terms of legislative history that “consistently characterizes the evil to be remedied—computer crime—as ‘trespass’ into computer systems or data”); *Black & Decker v. Smith*, 568 F. Supp. 939, 935 (W.D. Tenn. 2008) (“the legislative history supports the conclusion that Congress intended the CFAA to do ‘for computers what trespass and burglary laws did for real property’”) (quoting Orin S. Kerr, *Cybercrime’s Scope: Interpreting “Access” and “Authorization” in Computer Misuse Statutes*, 78 N.Y.U. L. REV. 1596, 1617 (2003)).

²¹⁹ 593 U.S. ___, slip diss. op. at 5, 141 S. Ct. 1648, 1664–65 (2021) (Thomas, J., dissenting).

²²⁰ See, e.g., *Facebook, Inc. v. Power Ventures, Inc.*, 844 F.3d 1058, 1065 (9th Cir. 2016) (“The CFAA prohibits acts of computer trespass by those who are not authorized users or who exceed authorized use”); *Shamrock Foods Co. v. Gast*, 535 F. Supp. 962 (D. Ariz. 2008) (“The general purpose of the CFAA was to create a cause of action against computer hackers (e.g., electronic trespassers)”) (internal quotation marks and citations omitted).

²²¹ See *In re America Online, Inc.*, 168 F. Supp. 2d 1359, 1370 n.8 (S.D. Fla. 2001) (refusing to “import” the “‘cluster of ideas’ association with common law ‘trespass’” into the CFAA).

²²² See *United States v. Valle*, 807 F.3d 508, 525 (2d Cir. 2015).

²²³ S. Rep. No. 99-432, at 7 (1986), *reprinted in* 1986 U.S.C.C.A.N. 2479, 2485.

²²⁴ *Id.* at 9, *reprinted in* 1986 U.S.C.C.A.N. at 2488.

²²⁵ H.R. Rep. No. 99-612, at 5–6 (1986).

The basis for the trespass comparison lies in thinking of a computer as a virtual “place.” The Supreme Court’s decision in *Packingham v. North Carolina* showed this sort of thinking when it described “cyberspace” as one of “the most important places (in a spatial sense) for the exchange of views.”²²⁶ Under this reasoning, access to a virtual place without permission thus analogizes to intrusion into a physical space. As with any reasoning by analogy, however, cyberspace-as-place has its critics and the aptness of the metaphor has been the subject of scholarly debate.²²⁷

If a computer is like a place, the question remains as to what sort of place it is: intrusion into a private home, for example, carries different implications about norms and harms than does trespass into a commercial space or open land.²²⁸ The crimes of trespass and burglary both involve intrusion into a space; which (if either) of these is more like the CFAA depends on the nature of the space and the intent of the intruder. Criminal trespass typically involves knowingly entering or remaining on property without permission.²²⁹ The crime escalates to burglary if the offender enters a building with the intent to commit a felony therein.²³⁰ Burglary is thus distinguished from trespass primarily in that a burglar must enter a building (not merely property) and must do so intending to commit a crime other than the entry itself.²³¹ If the essence of an (a)(2) crime is the unauthorized access (or “entry”) itself, then trespass may be the most apt analogy; this would also comport with the idea that privacy or the sanctity of personal space is the primary interest at stake in both a trespass and (a)(2) crimes. But if the access-to-information element of (a)(2) is essential to the crime and similar enough to the felony intent element of burglary, then burglary might be a better analogy.

Brenner argues that the difference between cybercrime as trespass versus burglary is the difference between “hacking” and “cracking”: a “hacker,” who “does not intend to commit an offense or cause damage,” commits trespass, but a “cracker,” who “breaks into a computer or computer system with the purpose of committing an offense once inside,” is more like a

²²⁶ *Packingham v. North Carolina*, 582 U.S. ___, slip op. at 5 (2017).

²²⁷ See, e.g., Mark A. Lemley, *Place and Cyberspace*, 91 CALIF. L. REV. 521, 523 (2003) (arguing that “the cyberspace as place metaphor is not a particularly good one”); Julie E. Cohen, *Cyberspace as/and Space*, 107 COLUMBIA L. REV. 210, 211 (2007) (“What began as a relatively narrow critique of the property metaphor’s doctrinal and political entailments has now blossomed into a full-blown debate about the merits of cyberspatial reasoning and rhetoric”).

²²⁸ See Kerr, *Norms of Computer Trespass*, *supra* note 208, at 1150–52 (discussing how the norms of trespass vary by the nature of the space).

²²⁹ WAYNE R. LAFAVE, 3 SUBST. CRIM. L. § 21.2 (3d ed.); CHARLES E. TORCIA, 3 WHARTON’S CRIM. L. § 332 (15th ed.). See also Model Penal Code § 221.2.

²³⁰ LAFAVE § 21.1; TORCIA § 316. The crime was traditionally defined at common law as breaking and entering into the dwelling of another at night; modern statutes tend to dispose of the requirements that the entry be into a dwelling (as opposed to any building) or occur at night. See LAFAVE § 21.1(c).

²³¹ Common law also required “breaking and entering,” not mere trespass—the defendant must have created the opening through which they entered. Someone who came through an open door or window would not have been a burglar under common law. The requirement for a “breaking” for burglary has largely been abandoned, however. See LAFAVE § 21.1(a).

burglar.²³² The crime that the “cracker” might contemplate could be damaging the system or using the information for fraud or theft, for example—although any of those would seem to make the crime a potential (a)(4) or (a)(5) crime, not merely an (a)(2) offense.

As an example, consider the hypothetical scenario used in Chapter 3, in which a person finds a flaw in a website and uses it to download records. The analogy to trespass or burglary might depend on motive: If the person’s motive was merely to look around, his actions could be analogous to a trespass crime. But if the motive was profit, burglary could be the better comparison.

4.2.2. Empirical Analyses of Cybercrime

There have been few empirical analyses of CFAA sentencing. The earliest appears to be Anele Nwokoma’s 2000 report of Department of Justice statistics on CFAA charging and sentencing.²³³ Of the 50 cases in Nwokoma’s data set, the majority (27, or 54%) were charged under (a)(4). Section (a)(2)—which until 1996 was still limited to access to financial information²³⁴—was the second-most common charge, with 12 convictions (24%).²³⁵ The median amount of loss in those crimes was between \$10,000 and \$20,000.

In 2011, Marcum, Higgins, and Tewksbury analyzed how the demographics of convicted cybercriminals and the types of cybercrimes affected sentences, using data from the Department of Corrections in three western states.²³⁶ They found that female offenders received slightly longer sentences than male offenders, although the effect was minimal (but statistically significant at $p < 0.05$). They also found that cybercrime offenders convicted of fraud, identity theft, or destruction of property received longer sentences than offenders who had not been convicted of each of those crimes. Their article did not disclose a sample size, however, and may be biased due to having data only on offenders who were sentenced to prison or jail terms.

Ioana Vasiu and Lucian Vasiu published an analysis in 2014 of over 300 civil and criminal computer damage cases under 1030(a)(5).²³⁷ But although the authors describe their analysis as allowing “empirical categorization” of essential aspects of the cases, their article did not specify how they found or selected those cases and did not contain any quantitative analysis or summary statistics.

²³² Brenner, *supra* note 217, at 80.

²³³ See Anele Nwokoma, *Process Evaluation of the Computer Fraud and Abuse Act of 1986*, 17 PROC. INFO. SYS. EDUC. CONF. § 128 (2000).

²³⁴ The expansion of (a)(2) to cover unauthorized access to government computers and “protected computers” was added in 1996. See Economic Espionage Act of 1986, Pub. L. No. 104-294 § 201, 110 Stat. 3488, 3491–92 (1996).

²³⁵ Ten of those involved credit card fraud; the other two involved embezzlement from financial institutions.

²³⁶ Catherine D. Marcum, George E. Higgins & Richard Tewksbury, *Doing Time for Cyber Crime: An Examination of the Correlates of Sentence Length in the United States*, 5 INT’L J. OF CYBER CRIMINOLOGY 825 (2011).

²³⁷ Ioana Vasiu & Lucian Vasiu, *Break on Through: An Analysis of Computer Damage Cases*, 14 U. PITT. J. TECH. L. & POL’Y 158 (2014).

Perhaps the most comprehensive empirical look at cybercrime—and part of the inspiration for the work in this chapter—is Jonathan Mayer’s 2016 article *Cybercrime Litigation*.²³⁸ Mayer compiled a data set of 325 civil pleadings and 133 criminal defendants, from 1986 to about 2013, using Bloomberg Law and Westlaw keyword searches of court filings, supplemented by Department of Justice announcements. He also obtained aggregate data from the Department of Justice and the Syracuse TRACfed project.²³⁹ Among Meyer’s findings was that sentence lengths had increased over the years but the proportion of offenders who received sentences of incarceration declined. He also reported statistics on charging practices and fact patterns, the latter of which showed that “about half of the prosecutions did not involve technical circumvention of an access control.”²⁴⁰ One limitation in Meyer’s analysis, however, is the data collection method: it is not clear whether the 133 sentences in his data set obtained through a keyword search are representative. The work in this chapter tries to improve on Mayer’s work with a larger and more comprehensive data set.²⁴¹

Most recently, in 2019 the U.S. Sentencing Commission released a detailed report on 29 different types of economic crimes that are sentenced under § 2B1.1 of the guidelines.²⁴² One of those types was “computer-related fraud,” which included also the CAN-SPAM Act and the Stored Communications Act in addition to the CFAA.²⁴³ The report is therefore not strictly an analysis of the CFAA, even where it reports on computer crime. But the level of detail in which it dissects sentencing data from one year (2017) is exhaustive.

Matthew B. Kugler’s 2016 article appears to be the only empirical analysis of cybercrime attitudes to date (other than Chapter 3 of this thesis).²⁴⁴ Kugler presented a representatively-weighted sample of 593 participants with several scenarios in one of three categories: misuse of an employer’s computer, accessing a neighbor’s WiFi network, or accessing a business’s website in violation of its terms and conditions. Participants were asked about the extent to which the action was unauthorized, how morally blameworthy it was, and how (or if) the actor should be punished. Notably for the purposes of this work, Kugler asked participants to rate the appropriate punishment in terms of comparison to other crimes: that is, should the actor in the vignette be given a punishment equivalent to that for a parking ticket, a

²³⁸ Jonathan Mayer, *Cybercrime Litigation*, 164 U. PENN. L. REV. 1453 (2016).

²³⁹ *Id.* at 1472 n.88 (2016). See also Federal Criminal Case Processing Statistics, BUREAU OF JUSTICE STATISTICS, <https://www.bjs.gov/fjsrc/>

²⁴⁰ Mayer, *supra* n.238, at 1484. The categories Mayer used in his coding of “underlying conduct” were “Misappropriating Information,” “Accessing Another Person’s Account,” “Financial Misfeasance,” “Editing or Deleting Information,” “Malware,” “Software Disruption of a Computer System,” “Unspecified Breaking In,” and “Hijacking Another Person’s Account.”

²⁴¹ See *infra* Section 4.3.1.

²⁴² COURTNEY SEMISCH, U.S. SENTENCING COMMISSION, WHAT DOES FEDERAL ECONOMIC CRIME REALLY LOOK LIKE? (2019), <https://www.ussc.gov/research/research-reports/what-does-federal-economic-crime-really-look-like>.

²⁴³ See 18 U.S.C. § 1037 (CAN-SPAM Act); 18 U.S.C. § 2701–13 (Stored Communications Act); SEMISCH at 36.

²⁴⁴ See Matthew B. Kugler, *Measuring Computer Use Norms*, 84 GEO. WASH. L. REV. 1568 (2016).

petty theft, a burglary—or no punishment at all. Kugler’s results defy easy summary, but they show that people have “sophisticated and nuanced views of what constitutes appropriate and inappropriate computer use.” They also showed a surprising willingness among some participants to impose criminal penalties for what others might see as minor offenses (or non-offenses), such as using a neighbor’s unsecured wireless network without permission.

4.3. Analysis of 1030(a)(2), Trespass, Burglary, and Fraud Sentences

4.3.1. Data and Methodology

I collected data on CFAA sentences from two sources: (1) the United States Sentencing Commission’s data files for individual offenders (which I refer to as the “USSC data set”)²⁴⁵ and (2) case documents gathered from PACER,²⁴⁶ Bloomberg Law, and Westlaw (“PACER data”).

Sentencing Commission Data Files

The Sentencing Commission’s data files include information on all individual offenders convicted of federal felonies and Class A misdemeanors.²⁴⁷ The data set does not include information on organizational defendants, cases in which all charges were dismissed or the defendant was acquitted of all charges, or cases in which all charges were petty offenses.²⁴⁸ Variables in the Commission’s data files describe offender characteristics (e.g., month and year of birth, gender, race, criminal history category), sentence information (e.g., statutes of conviction²⁴⁹ and sentences imposed), and detailed guideline calculation factors (e.g., which guideline sections were applied and how the default offense level was adjusted based on particular facts of the case). The data set excludes case identifiers such as birthdates, offender names, sentencing dates, and docket numbers.²⁵⁰

I downloaded Commission data files for fiscal years 2005 through 2019. I removed cases from the FY 2005 data that were decided before the Supreme Court’s decision in *Booker*.²⁵¹ I

²⁴⁵ <https://www.ussc.gov/research/datafiles/commission-datafiles>

²⁴⁶ <https://pacer.uscourts.gov/>

²⁴⁷ U.S. SENTENCING COMM’N, VARIABLE CODEBOOK FOR INDIVIDUAL OFFENDERS: STANDARDIZED RESEARCH DATA DOCUMENTATION FOR FISCAL YEARS 1999–2020 (2021), https://www.ussc.gov/sites/default/files/pdf/research-and-publications/datafiles/USSC_Public_Release_Codebook_FY99_FY20.pdf; CHRISTINE KITCHENS, U.S. SENTENCING COMM’N, INTRODUCTION TO THE COLLECTION OF INDIVIDUAL OFFENDER DATA BY THE UNITED STATES SENTENCING COMMISSION 2 (2009).

²⁴⁸ KITCHENS at 2.

²⁴⁹ The USSC data files encode statutes as an unpunctuated string (e.g., 18 U.S.C. 1030(a)(2) would appear in the data set as “181030A2”), which in some cases made distinguishing between statutes impossible. For example, the data set encodes both 18 U.S.C. § 1028(a) and 18 U.S.C. § 1028A as “181028A.” Also, not all records included subsections (see Section 4.3.2.2 and Table 11, *infra*).

²⁵⁰ Data files up to FY2005 included dates of birth. Subsequent data files list only the month and year of birth. Data files contained sentencing dates through FY2004, after which the data files contain only the month and year of sentencing.

²⁵¹ *United States v. Booker*, 543 U.S. 220 (2005).

also omitted partial-year data for calendar year 2019. The portion of the USSC data set that I downloaded includes 1,060,213 post-*Booker* criminal sentences imposed from January 2005 through December 2018, including 1,209 CFAA sentences.

Court Filings (“PACER data”)

The second data set is based on information collected from 2,554 court dockets and filings for 1,734 criminal defendants. To locate these documents, I generated a list of CFAA cases by searching each federal district court’s PACER site. I used the Criminal Cases Report function to search for all pending or terminated cases filed between January 1, 2000, and October 18, 2018 (the date of the search) with a citation of “18:1030A.F” or “18:1030A.M.”²⁵² I reviewed the dockets of each case to determine whether each resulted in a final conviction for a CFAA charge. From the case documents, I recorded the case disposition (plea, guilty verdict, acquittal, etc.), number of CFAA counts of conviction by subsection, sentencing date, and sentence. If 1030(a)(2) was among any of the statutes of conviction, I also recorded a summary of the facts of the case based on plea agreements, indictments or informations, complaints, or sentencing memoranda.

During this process, I preferred plea documents to indictments and complaints. If a plea agreement (or associated factual statement) contained a detailed stipulation of facts, I did not also retrieve a complaint or indictment. But if the plea agreement only recited the elements of the crime or contained only a superficial description of the facts, I retrieved other documents to fill in more details. In a few cases where none of these documents described the facts of a case, I relied on press releases or news reports, if I could find them. I did not retrieve all these documents unless necessary: if a more authoritative document (e.g., a plea agreement) explained the facts of a case, I did not retrieve less-authoritative documents (e.g., a complaint) for that case.

After entering the data from case documents, I coded (a)(2) sentences according to the general fact pattern and the type of access. The method used for the coding was based on observation of common fact patterns in the data (i.e., I did not attempt to tie the categorization into existing taxonomies, nor did I attempt to create my own taxonomy). I created two variables: *Fact Code* and *Access Type*. The *Fact Code* coded the general fact pattern. *Access Type* coded whether the access was “unauthorized” or “exceeded authorized access.” I coded conduct as unauthorized if the offender did not have permission to use the system or the data they accessed. I coded conduct as exceeding authorized access if the offender had permission to use the system and the data on that system but used that system or data for an unauthorized purpose. I chose these codings because they partition the cases in the data set rather neatly.

²⁵² For N.D. Georgia, which has no Criminal Cases Report function, I searched Bloomberg for cases matching “18:1030,” inspected the results manually, and added CFAA sentences to the data set.

After eliminating duplicates,²⁵³ cases in which all CFAA charges were dismissed, and other cases that did not result in a conviction and sentence, the remaining data consisted of 1,143 CFAA sentences imposed between January 31, 2005, and December 17, 2018.²⁵⁴

Merged Data Set

I combined the USSC and PACER data sets into a single data set containing the facts, offender characteristics, and sentencing calculations of 1,095 CFAA sentences imposed between January 31, 2005, and December 17, 2018. I combined the data sets by comparing information that appeared in both. Fields that occurred in both data sets included the month and year of sentencing; the terms of imprisonment, supervised release, or probation; the amounts assessed as fine, restitution, and special assessment; and the federal district court that imposed the sentence.²⁵⁵ These common fields were sufficient to cross-reference almost all the records in the two data sets. The 1,095 cross-matched records represent 96% of the 1,143 sentences in the PACER data set for January 31, 2005, through December 17, 2018, and 91% of the 1,209 CFAA sentences in the USSC data set for that time period.²⁵⁶

I also used the PACER data to update and correct the USSC data set. There were a few records that clearly matched between data sets except for an error or omission in the USSC data. For example, the USSC data files sometimes lacked restitution amounts that were available in court documents.²⁵⁷ The PACER data also provided CFAA subsection information for 40 sentences in the USSC data set that did not include that information.

²⁵³ Because I searched for felony and misdemeanor cases separately, some cases appeared in both searches. Some cases that were transferred between jurisdictions also appeared in the dataset twice.

²⁵⁴ The PACER search also turned up 13 cases with CFAA charges in which the judgments were sealed. I do not include these cases in the results.

²⁵⁵ Most of the cases in which matching records between data sets was at all difficult were where a single case had several defendants all of whom received the same sentence. In *United States v. Collins*, No. 11-cr-00471 (N.D. Cal. July 13, 2011), twelve defendants convicted of a single count each of violating 1030(a)(5) each received sentences of one year probation, \$5,600 restitution, and a \$25 special assessment. A thirteenth defendant, convicted under two (a)(5) counts, received the same sentence except that the special assessment was \$50.

²⁵⁶ Unmatched sentences were not evenly distributed across districts. Districts with five or more unmatched sentences included TX-N (12 unmatched sentences, which represent 37.5% of the 32 CFAA sentences for TX-N in the USSC data set), MO-W (11, 55% of 20), CA-E (8, 18.6% of 43), and CA-C (7, 7.5% of 93). Certain fact patterns were also overrepresented in unmatched sentences—for example, of the 48 cases in the PACER data for which I could not find a corresponding sentence in the USSC data set, 11 were IRS agents who used their access to taxpayer information for unauthorized purposes (out of 49 sentences with that fact pattern).

²⁵⁷ The court documents containing restitution information may have been filed after the sentencing information was provided to the Commission.

4.3.2. What do 1030(a)(2) Computer Crimes Look Like?

4.3.2.1. Fact Patterns

Section 1030(a)(2) covers a wide range of conduct, from high-profile “hacking” cases to more mundane examples of employees accessing data outside of the scope of their employment agreements. Figure 7 illustrates the distribution of fact patterns in the PACER data set.

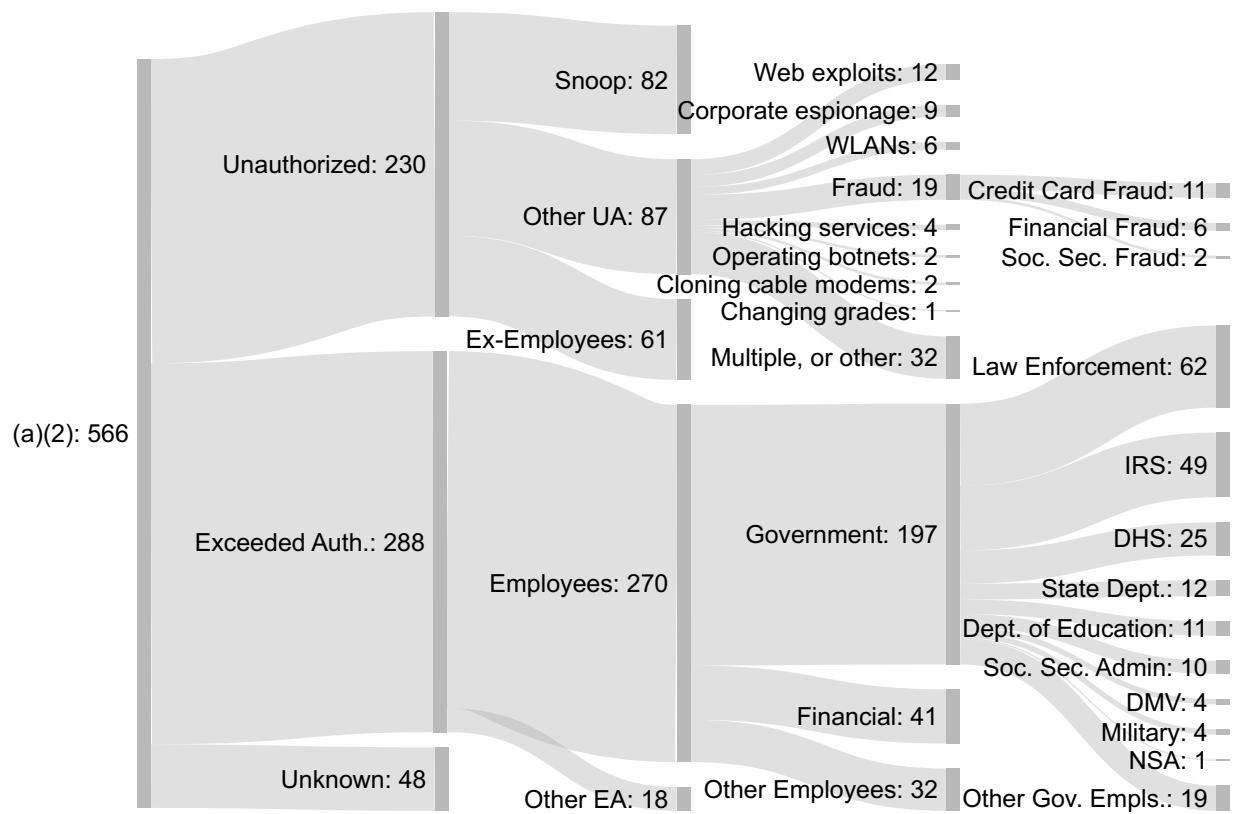


Figure 7: CFAA (a)(2) fact patterns in the PACER data set

Slightly more than half of the 563 (a)(2) sentences in the PACER data set involved access in excess of authorization (285, 50.6%), while unauthorized access comprised 40.9% (230) of (a)(2) sentences. I was unable to determine the access type for the other 48 (8.5%). Considering only the 421 (a)(2) sentences in which the CFAA was the only statute of conviction, 221 (52.5%) were in exceeding-access cases, 161 (38.2%) involved unauthorized access, and I was unable to determine the type of access for 9.3% (39).

Nearly half of (a)(2) cases (270, 47.7%) involved employees using their authorized access for unauthorized purposes. These included 197 sentences for government employees (of whom there were 62 in law enforcement, 49 in the IRS, 25 in DHS, and 61 in various other areas of government), 41 for employees of financial institutions, and 32 for other employees (or employees of unknown types of employers). The other 18 sentences for exceeding authorized

access included 15 sentences for violating terms of service, 2 insider-trading sentences,²⁵⁸ and one sentence for which I could find no information other than charging documents alleging that the defendant “exceeded authorized access.”²⁵⁹

The terms-of-service sentences involved a variety of fact patterns. For example, the defendant in *United States v. Lowson* pled guilty to writing code to purchase tickets in bulk online from Ticketmaster for resale to ticket brokers; the code bypassed Captcha challenges and other methods meant to ensure that ticket brokers were people.²⁶⁰ Some terms-of-service sentences were for resellers who used ordering systems to purchase equipment at discounts for which they were not authorized.²⁶¹ In another notable case, the CFAA charge was based on allegations that the defendant had obtained employment as a contractor by “misrepresent[ing] the extent of his employment background.”²⁶² Other terms-of-use cases included a defendant who made insider trades in violation of a user agreement that authorized usage of the account only for lawful purposes;²⁶³ a person who took professional licensing exams in other people’s names for \$1500 to \$2000 per test in violation of the testing center’s terms of use;²⁶⁴ and another who ran queries in the National Loan Student Data system to generate leads for loan consolidation in violation of the data system’s use agreement.²⁶⁵ There were also a handful of financial or fraud crimes where violating of terms of service was the hook for an (a)(2) charge.²⁶⁶

²⁵⁸ See Information, *United States v. Reier*, No. 04-cr-00583 (D.N.J. June 27, 2007); Information, *United States v. Sacks*, No. 04-cr-00583 (D.N.J. Nov. 13, 2007). Both defendants had been charged with securities fraud under 15 U.S.C. § 78j(b), 17 U.S.C. § 240.10b-5, and 18 U.S.C. § 2 before pleading to one count each of exceeding authorized access under § 1030(a)(2). I was unable to find any court documents explaining how the defendants exceeded authorized access by purchasing stock on an insider trading tip. It is possible (though this is pure speculation) that they violated their brokers’ terms of service or that they made the stock purchases on an employer’s computer.

²⁵⁹ Criminal Information, *United States v. Gaskins*, No. 05-mj-00207 (E.D.N.C. Oct. 25, 2005).

²⁶⁰ Superseding Indictment at 1–3, *United States v. Lowson*, No. 20-cr-00114 (D.N.J. Apr. 20, 2010). See also Kim Zetter, *Wiseguys Plead Guilty in Ticketmaster Captcha Case*, WIRED (Nov. 19, 2010), <https://www.wired.com/2010/11/wiseguys-plead-guilty/>.

²⁶¹ See, e.g., Superseding Information at 1–2, *United States v. Maldonado*, No. 12-cr-00076 (C.D. Cal. July 25, 2016) (purchasing Cisco equipment using a discount code the purchaser was not authorized to use for the purpose of the purchase); Plea Agreement at 7–8, *United States v. Ashraf*, No. 13-cr-00088 (C.D. Cal. Nov. 16, 2015) (using an online ordering system to purchase Hewlett Packard products for “unauthorized end users”).

²⁶² See Defendant’s Sentencing Memorandum at 6, *United States v. Parker*, No. 12-cr-00059 (E.D. Va. May 16, 2012). The defendant also pled guilty to five counts of access device fraud, aggravated identity theft, and wire fraud, so the 24-month sentence on the CFAA charge had no real effect on the concurrent 63-month sentence that was imposed for the other charges. See Docket, *id.*

²⁶³ See Bill of Information at 1–2, *United States v. Mead*, No. 05-cr-00066 (W.D.N.C. Mar. 22, 2005).

²⁶⁴ See Statement of Facts, *United States v. Thai*, No. 12-cr-00065 (E.D. Va. Feb. 24, 2012).

²⁶⁵ See Plea Agreement at 14–17, *United States v. Breidert*, No. 09-cr-00290 (M.D. Fla. June 23, 2009).

²⁶⁶ See, e.g., Plea Agreement at 11, *United States v. Williams*, No. 09-cr-00298 (M.D. Fla. June 26, 2009) (making Western Union wire transfers to receive the proceeds of credit card fraud in violation of Western Union’s terms of service); Indictment at 2–3, Plea Agreement at 1–2, and Docket, *United States v. Obaid*, No. 15-cr-20040 (D. Kan. 2016) (“exceeding authorized access” by accessing card websites to redeem reward points obtained in a scheme to obtain those points by making purchases then canceling the purchases).

The data set does not include any criminal sentences for “web scraping,” which has been the subject of much scholarly interest and civil litigation.²⁶⁷ The closest thing to a scraping case in the data set is *United States v. Lowson*,²⁶⁸ but the violation of the terms of service at issue in *Lowson* included the purchase and resale of tickets, which went beyond the mere automated collection of publicly available information typical of web scraping.²⁶⁹ The lack of a criminal sentence for web scraping, at least among the (a)(2) sentences in my data set, suggests that although civil litigation over web scraping has been common, criminal prosecutions are rare.²⁷⁰

Among cases involving unauthorized access, employees—or, more precisely, former employees—also featured prominently. This fact pattern accounted for 61 sentences (11%). These cases generally involved disgruntled ex-employees who accessed their former employers’ computer systems.

The other major category of unauthorized access cases, with 82 sentences in the PACER data set (14%), is what I broadly term “snooping”—accessing a system to get private information. These are access-to-information cases at which 1030(a)(2) was squarely aimed. These cases include defendants who installed spyware, accessed e-mail or social media accounts, downloaded photos or videos from cloud accounts without authorization, or engaged in cyberstalking.

The 87 remaining unauthorized-access sentences (15%) represent a grab bag of computer-related misconduct, some of which can still reasonably be called “hacking,” some of which cannot. Of these 87, 19 involved some form of fraud: identity fraud (2 sentences), credit-card fraud (11 sentences), or financial fraud (using the victim’s credentials to steal money from bank accounts) (6 sentences). Twelve involved web exploits—defendants who hacked into a site and either downloaded information to sell it or hoped that the site owners would pay to avoid

²⁶⁷ See, e.g., *Hi-Q Labs, Inc. v. LinkedIn Corp.*, 978 F.3d 985 (9th Cir. 2017), *vacated*, ___ S. Ct. ___ (June 14, 2021); *EF Cultural Travel BV v. Explorica, Inc.*, 274 F.3d 577 (1st Cir. 2001), *abrogated by* *Van Buren v. United States*, 593 U.S. ___ (2021); *Sandvig v. Barr*, 451 F. Supp. 3d 73 (D.D.C. 2020); *QVC, Inc. v. Resultly, LLC*, 159 F. Supp. 3d 576 (E.D. Pa. 2016); *Craigslist Inc. v. 3Taps Inc.*, 942 F. Supp. 2d 962, 968–70 (N.D. Cal. 2013); Sara R. Benson, *Social Media Researchers and Terms of Service: Are We Complying with the Law?* 47 AIPLA Q.J. 191 (2019); Christine D. Galbraith, *Access Denied: Improper Use of the Computer Fraud and Abuse Act to Control Information on Publicly Accessible Internet Websites*, 63 MD. L. REV. 320 (2004); Zachary Gold & Mark Latonero, *Robots Welcome? Ethical and Legal Considerations for Web Crawling and Scraping*, 13 WASH. J.L. TECH. & ARTS 275, 295–99 (2018); Michael J. O’Connor, *The Common Law of Cyber-Trespass*, 85 BROOK. L. REV. 421, 460–62 (2020); Andrew Sellars, *Twenty Years of Web Scraping and the Computer Fraud and Abuse Act*, 24 B.U. J. SCI. & TECH. L. 372 (2018); Benjamin L.W. Sobel, *A New Common Law of Web Scraping*, 25 LEWIS & CLARK L. REV. 147 (2021); Jamie L. Williams, *Automation Is not “Hacking”: Why Courts Must Reject Attempts to Use the CFAA as an Anti-Competitive Sword*, 24 B.U. J. SCI. & TECH. L. 416 (2018).

²⁶⁸ See *supra* note 260 and accompanying text.

²⁶⁹ The definition of “web scraping” is itself surprisingly nuanced. See Sellars, *supra* note 267, at 381–88.

²⁷⁰ Sellars lists a few “famous” prosecutions of scraping under the CFAA, including the cases against Aaron Swartz, No. 11-cr-10260 (D. Mass. filed July 14, 2011), and Andrew Auerenheimer. 748 F.3d 525 (3d Cir. 2014). See Sellars, *supra* note 267, at 377 n.40. Neither case resulted in a sentence; Swartz’s case ended when he killed himself and Auerenheimer’s case was dismissed on appeal for improper venue. See 748 F.3d at 541.

having that information or the vulnerability released.²⁷¹ Nine sentences were the result of corporate espionage.²⁷² Defendants in six cases had used wireless LANs without permission.²⁷³ Four sentences involved defendants who offered hacking services like “needapassword.com” or “hackthissite.org”; this category also includes Ross Ulbricht, who ran the Silk Road black market site that traded in, among other things, hacking tools. Two sentences were for defendants who ran botnets. Another two defendants had sold cloned cable modems. One defendant even had the classic *WarGames* fact pattern of hacking into the school’s system to change grades.²⁷⁴ The

²⁷¹ See, e.g., Plea Agreement at 8, *United States v. Chowdry*, No. 10-cr-00264 (D. Ariz. Aug. 8, 2010) (discovering a vulnerability in a GoDaddy website and using it to download information from 6,000 accounts); Factual Basis, *United States v. Whitaker*, No. 14-cr-00420 (M.D.N.C. Feb. 23, 2015) (exploiting a web vulnerability in a credit union’s website without accessing financial records, and downloading about 64,000 records from a .mil domain via a website vulnerability); Plea Agreement at 5, *United States v. Axelrod*, No. 14-cr-01209 (D. Ariz. Dec. 16, 2014) (using a vulnerability the defendant discovered to access, but not download, files on a community college’s server). Perhaps the most egregious case in this category was that of Ardit Ferizi, who “gained access” to a company’s web server, culled information on 1,300 U.S. civilian and military employees, and sent that information to ISIL. Statement of Facts, *United States v. Ferizi*, No. 16-cr-00042 (E.D. Va. June 15, 2016).

²⁷² For example, the defendant in one case used credentials supplied by the former employee of a competitor to access the competitor’s database and view customer information. See Information, *United States v. Wolf*, No. 09-cr-00463 (N.D. Cal. Apr. 30, 2009). In another, the president of a security consulting company looked for vulnerabilities and exploited them to retrieve documents, which they hoped would create publicity and bring in new clients. See Indictment, *United States v. O’Keefe*, No. 03-cr-02659 (S.D. Cal. Sept. 23, 2003). Another defendant accessed approximately 1300 accounts at competing insurers using customer SSNs and birth dates to try to get the customers to transfer their accounts. See Plea Agreement, *United States v. Berger*, No. 09-cr-00066 (D. Conn. Mar. 24, 2009). And in the world of sports, a baseball executive for the St. Louis Cardinals received a 46-month sentence and paid almost \$280,000 in restitution for using the passwords of former Cardinals employees to view confidential information on the Houston Astros’ internal web system. See Plea Agreement, *United States v. Correa*, No. 15-cr-00679 (S.D. Tex. Jan. 8, 2016).

²⁷³ In only one of these cases was the CFAA the sole count of conviction. In that case, the defendant had used an open wireless network to do a Google search while stalking and then harassing a woman. The defendant was charged with making interstate threats, 18 U.S.C. §§ 875(d), 2261(A), but pled down to a single charge of computer fraud. See Sentencing Memorandum of the United States, *United States v. Klig*, No. 09-cr-00856 (S.D.N.Y. Sept. 24, 2010), and the docket in the same case.

The CFAA charges in the other five cases were among multiple charges in cases in which prosecutors had “thrown the book” at the defendants. One case, like *Klig*, involved harassment: Barry Ardolf had used a neighbor’s wireless LAN to send spoofed e-mail containing child porn and to send other spoofed e-mails with threats to the Vice President, Governor, and a senator, all as part of a harassment campaign against Ardolf’s neighbors. See *United States v. Ardolf*, 683 F.3d 894, 897–98 (8th Cir. 2012). In three other cases, the defendants had accessed wireless networks to download child pornography. See Stipulated Factual Basis, *United States v. Vandiver*, No. 13-cr-00046 (S.D. Ind. Nov. 22, 2013) (“crack[ing]” a wireless LAN); Plea Agreement at 16-18, *United States v. Johnson*, No. 11-cr-00287 (E.D. Wis. Jul 25, 2012) (hacking into an encrypted WLAN); Information at 1–2, *United States v. Duncan*, No. 12-cr-00006 (S.D. Ohio Jan. 5, 2012) (downloading child pornography using the unsecured wireless LAN of a hotel at which the defendant was not a customer). The remaining defendant had hacked into an encrypted wireless network to send a bomb threat to a mall. See Superseding Indictment at 2–3, *United States v. Barnhouse*, No. 13-cr-00659 (E.D. Pa. Jan. 30, 2014).

²⁷⁴ See Plea Statement at 4–7, *United States v. Li*, No. 06-cr-00081 (D. Utah Oct. 4, 2006) (hacking professor’s password to access a grading spreadsheet in an attempt to change the student’s grade). See also *WARGAMES* (United Artists 1983). (No, really, see it. It’s great.) The more serious scenario in that film—that of a teenager nearly launching global thermonuclear war by hacking into NORAD computers—was fresh in the public’s consciousness when the CFAA was enacted and was frequently cited at the time as the kind of threat the CFAA was needed to prevent. See, e.g., Decian McCullagh, *From ‘WarGames’ to Aaron Swartz: How U.S. Anti-Hacking Law*

other 32 unauthorized access cases involved multiple types of fact patterns or did not fit into the categories above.

4.3.2.2. CFAA Convictions by Subsection

The vast majority of CFAA convictions have been under subsections (a)(2), (a)(4), and (a)(5), with (a)(2) representing nearly half of all CFAA convictions in the data set (572/1,209). Together, these three subsections account for 95% of all CFAA sentences from 2005 through 2018 (1,146/1,209). Sentences of conviction under (a)(1), (a)(3), (a)(6), and (a)(7) are much rarer: the 48 sentences under these subsections represent only 4% of CFAA sentences in that time.²⁷⁵

Table 11: CFAA convictions by subsection, 1/31/05– 12/31/18

<i>Subsection</i>	<i>Description</i>	<i>N</i>	<i>%</i>
(a)(1)	Obtaining national security information	4	0.3
(a)(2)	Accessing a computer and obtaining information	572	47.3
(a)(3)	Accessing a government computer	10	0.8
(a)(4)	Accessing a computer for fraud	242	20.0
(a)(5)	Damaging a computer	355	29.3
(a)(6)	Trafficking in passwords	20	1.7
(a)(7)	Computer threats and extortion	14	1.2
(b)	Attempt and conspiracy	84	6.9
Unknown	Subsection unknown	9	0.7
All		1,209	100.0

Note: Counts do not add up to the total because of convictions that included multiple CFAA subsections (30 sentences included multiple 1030(a) subsections; 1030(b) appeared with 1030(a) subsections 68 times). “Unknown” means no subsection was specified in the USSC dataset or in the PACER files I reviewed. Figures for 1030(b) exclude sentences in which a 1030(a) subsection was also a statute of conviction. Source: USSC and PACER combined data set.

The frequency with which (a)(2) appears should not be surprising given the broad scope of conduct covered by (a)(2). A violation requires only that the defendant either (1) accessed a computer without authorization, or (2) accessed a computer in excess of authorization and then obtained information the defendant was “not entitled so to obtain.”²⁷⁶ The information need not be sensitive or extensive. In fact, it would be difficult to log into a computer without obtaining *some* information.

Subsection (a)(2) may also be an option for prosecutors in cases where (a)(4) might seem a better fit for the crime. In contrast to the broad elements of (a)(2), (a)(4) requires showing that the defendant had an intent to defraud, that the computer access furthered that fraud, and that the

Went Astray, C|NET (Mar. 13, 2013), <https://www.cnet.com/news/from-wargames-to-aaron-swartz-how-u-s-anti-hacking-law-went-astray/>.

²⁷⁵ Six of those sentences were all from the same district and the same year: the Eastern District of Virginia in 2010.

²⁷⁶ 18 U.S.C. §§ 1030(a)(2), (e)(5); *Van Buren v. United States*, 593 U.S. ___, slip op. at 13 (2021).

defendant obtained something of value.²⁷⁷ The latter statute might seem to offer a higher maximum sentence—subsection (a)(4) has a baseline maximum sentence of five years, versus one year under (a)(2)—but the maximum sentence under (a)(2) is also five years if the offense was “committed for purposes of commercial advantage or private financial gain,” “committed in furtherance of any criminal or tortious act,” or if the total value of the information obtained was more than \$5000.²⁷⁸ There is therefore substantial overlap between the elements of (a)(4) and the showings required for a five-year maximum under (a)(2): using a computer to commit fraud is likely to be both “for purposes of commercial advantage or private financial gain” and “in furtherance of [a] criminal or tortious act.”²⁷⁹ And many forms of fraud using computers will also involve obtaining information like bank records, social security numbers, or credit card numbers. Data from court filings shows many (a)(2) sentences that seem like they could have been (a)(4) sentences. Given this overlap, it is somewhat surprising that (a)(2) and (a)(4) are co-statutes of conviction extremely rarely. Only 5 of the 1,134 CFAA sentences in the database included both (a)(2) and (a)(4) as statutes of conviction.

The less frequently used subsections (a)(1), (a)(3), and (a)(6) cover conduct that can be charged under statutes that are more attractive to prosecutors. Improper access to national security information, (a)(1), is more commonly charged under 18 U.S.C. § 793(e), “for which guidance and precedent are more prevalent”²⁸⁰ and which has the same ten-year maximum sentence as 1030(a)(1). Access to a government computer under 1030(a)(3) offers little or no advantage over (a)(2) from a prosecutor’s perspective because it has a one-year maximum sentence without the aggravating factors that can turn an (a)(2) violation into a felony with a five-year maximum sentence.²⁸¹ Trafficking in passwords, (a)(6), can usually be charged under 18 U.S.C. § 1029, which has a higher maximum sentence of ten years for a first offense²⁸² compared to the one-year maximum for a first offense under 1030(a)(6).²⁸³

It is not immediately clear why 1030(a)(7) is not more frequently charged. The subsection was added to the CFAA to “fill perceived gaps in . . . existing anti-extortion statutes.”²⁸⁴ Since 2005, however, 1030(a)(7) appears as a statute of conviction only 14 times.²⁸⁵ By contrast, 18 U.S.C. § 875(d) (interstate communication threatening to injure property)

²⁷⁷ 18 U.S.C. § 1030(a)(4).

²⁷⁸ 18 U.S.C. § 1030(c)(2)(B), (c)(3).

²⁷⁹ Access under (a)(2) must be intentional, however, while (a)(4)’s standard is “knowingly and with intent to defraud.” See *United States v. Nosal*, 844 F.3d 1024, 1033 (9th Cir. 2016) (“Subsection 1030(a)(2) . . . mirrors (a)(4) but requires that access be intentional”).

²⁸⁰ DEP’T OF JUSTICE, PROSECUTING COMPUTER CRIMES 15, <https://www.justice.gov/sites/default/files/criminal-ccips/legacy/2015/01/14/ccmanual.pdf> [hereinafter DOJ COMPUTER CRIMES MANUAL].

²⁸¹ 18 U.S.C. § 1030(c)(2)(A), (B); *See also* DOJ COMPUTER CRIMES MANUAL, *supra* note 280, at 23.

²⁸² 18 U.S.C. § 1029(c)(1).

²⁸³ 18 U.S.C. § 1030(c)(2)(A).

²⁸⁴ DOJ COMPUTER CRIMES MANUAL, *supra* note 280, at 55.

²⁸⁵ Combined USSC and PACER data set. The USSC data set shows 13 convictions under (a)(7); the other was found through the PACER data.

appears 141 times and 18 U.S.C. § 1951 (interference with commerce by extortion) occurs 10,672 times.²⁸⁶ Section 1951, although it does not overlap perfectly with 1030(a)(7),²⁸⁷ has a higher maximum sentence for a first offense (twenty years versus five) and is probably more familiar to prosecutors. Or it may simply be that there are not many (a)(7) violations that do not also violate some other CFAA subsection. For example, a defendant who finds a flaw in a web site and threatens to make that flaw public unless he is paid could also be charged under (a)(2).²⁸⁸

4.3.2.3. CFAA Convictions over Time

As shown in Figure 8, the number of CFAA sentences has declined over the years, from a high of 126 sentences in 2009 to a low of 54 in 2017. There was an uptick in 2018, however, when 63 CFAA sentences were imposed.

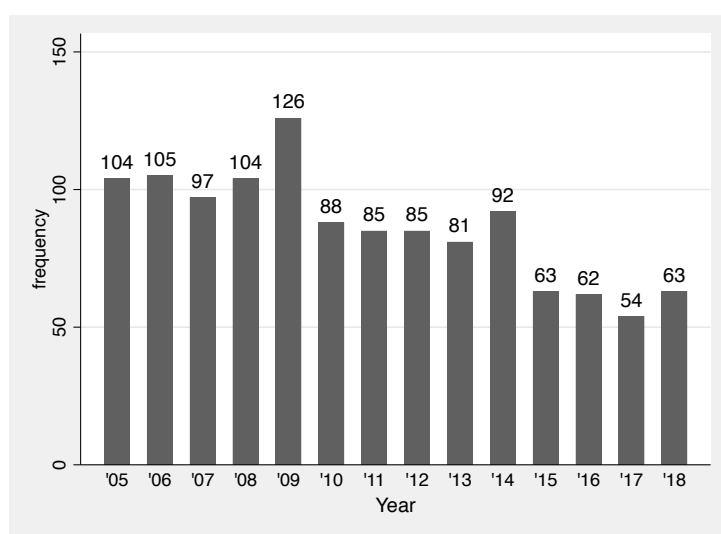


Figure 8: Number of CFAA sentences by year, 1/31/05– 12/31/18

Note: “Year” indicates the calendar year in which each sentence was imposed. Source: USSC and PACER combined data set.

²⁸⁶ USSC data set.

²⁸⁷ A violation of 18 U.S.C. § 1951 requires that the defendant obstructed, delayed, or affected commerce by extortion or threat of physical violence to person or property.

²⁸⁸ In *United States v. Potere*, for example, a law firm associate downloaded confidential documents and threatened to send them to a legal blog unless the firm paid him \$210,000 and a piece of artwork. He pled guilty to an (a)(2) charge and received a five month jail sentence. Plea Agreement, *United States v. Potere*, No. 17-cr-00446 (C.D. Cal. Oct. 18, 2017); *see also* the judgment and docket in the case. In *United States v. Morlock*, the defendant pled guilty to an (a)(2) charge for hacking into an internet retailer’s website, downloading order histories and a credit card database, then threatening to disclose that information unless the retailer paid him \$80,000. Government’s Sentencing Memorandum, No. 04-cr-572 (E.D.N.Y. July 22, 2013). In *United States v. Mengler*, No. 08-cr-3203, the defendant found a flaw in a Maserati dealer’s web site, exploited that flaw to obtain the names and addresses of people who had received a promotional mailer from the dealer, then threatened to publicize the security flaw and release the names and addresses unless the dealer paid him. Prosecutors charged Mengler with one count under 18 U.S.C. § 1030(a)(5) and four counts under 18 U.S.C. § 875(d). *See* Indictment, *United States v. Mengler*, No. 08-cr-3203 (S.D. Cal. Sept. 18, 2008). Mengler pled down to a single (a)(2) count and paid a \$1000 fine (plus three years unsupervised probation and a \$100 assessment).

Table 12: Number of 1030(a)(2) sentences by access type and year

<i>Year</i>	<i>Exceeded</i>	<i>Unauth.</i>	<i>Unknown</i>	<i>Total</i>
2005	19	5	12	36
2006	26	16	6	48
2007	25	14	5	44
2008	29	9	5	43
2009	42	22	4	68
2010	31	14	0	45
2011	28	16	1	45
2012	26	20	5	51
2013	20	18	2	40
2014	15	21	1	37
2015	7	20	3	30
2016	10	18	2	30
2017	5	19	2	25
2018	5	19	0	24
Total	288	230	48	566

Note: This table shows the number of 1030(a)(2) sentences imposed in each calendar year according to whether access was in excess of authorization, was unauthorized, or could not be determined from court filings. Source: PACER data set.

Table 12 suggests that a possible explanation for the decline in CFAA cases over time is a decrease in CFAA convictions for exceeding-authorized-access cases. For 1030(a)(2) convictions (the only ones I analyzed for access type), the number of unauthorized-access convictions has remained fairly constant or even increased slightly over the past decade. But exceeding-authorized-access (a)(2) convictions have declined sharply over the same period. From 2005 through 2012, 66% of 1030(a)(2) of sentences with known fact patterns (226 of 342) were for using systems for unauthorized purposes. From 2013 through 2018, only 35% (62 of 176) were for exceeding authorized access.

4.3.2.4. CFAA Sentencing

4.3.2.4.1. 1030(a)(2) Sentences Compared to Other Subsections

Table 13 summarizes CFAA sentencing by subsection. It shows the frequency, for sentences in which the CFAA was the only substantive statute of conviction, with which CFAA sentences include prison time; the median number of months of jail or prison time for sentences that include prison; and the percentage of sentences that include orders to pay restitution. To avoid conflating factors from other statutes of conviction, the table (and Tables 14 and 15 below) shows only sentences in which the CFAA was the sole substantive count of conviction.²⁸⁹

²⁸⁹ Of the 1,209 CFAA sentences in the combined data set, 895 (74%) had only CFAA charges, as did 422 of the 572 (a)(2) sentences (74%). Non-substantive counts of conviction are 18 U.S.C. § 2 (principals), § 371 (conspiracy), and § 3571 (sentence of fine). The table does not exclude CFAA convictions that occurred along with convictions for petty offenses.

Table 13: CFAA Sentencing by subsection

<i>Access Type</i>	<i>Prison</i>	<i>Median Months</i>	<i>Restitution</i>	<i>N</i>
(a)(1)	0.0%	-	0.0%	2
(a)(2)	24.0%	9	33.7%	416
(a)(3)	0.0%	-	50.0%	6
(a)(4)	56.3%	12	77.0%	126
(a)(5)	47.0%	15	83.8%	296
(a)(6)	13.3%	5	73.3%	15
(a)(7)	87.5%	24	37.5%	8
(b)	66.7%	18	41.7%	12
Multiple	46.2%	11	76.9%	13
Unknown	100.0%	36	100.0%	1
Total	37.3%	12	57.7%	895

Note: “Prison” lists the percentage of sentences that included a term of imprisonment. “Median months” lists the median length of imprisonment for sentences that included any prison term. “Restitution” indicates the percentage of sentences that included restitution. This table excludes sentences with non-CFAA statutes of conviction (except 18 U.S.C. § 2 (principals), § 371 (conspiracy), § 3571 (sentence of fine), or petty offenses). Source: USSC data set supplemented by PACER data.

Sentences under 1030(a)(2) were less likely to involve prison time (24% of 422 vs. 49% of 473, $p < 0.001$, χ^2)²⁹⁰ or restitution (34% vs. 79%, $p < 0.001$, χ^2) than those for other CFAA crimes. The median sentence for (a)(2) crimes was also shorter than for other sections, at 8.5 months versus 15 months.

4.3.2.4.2. 1030(a)(2) Sentencing by Access Type and Fact Pattern

Table 14 breaks down sentencing by access type for 1030(a)(2) sentences in which CFAA charges were the only substantive charges of conviction.

Sentences for exceeding-authorized-access (a)(2) cases received lighter sentences overall than unauthorized access cases. The unauthorized-access convictions were more likely to include prison time, had longer median sentences, and were more likely to require restitution.

Table 14: 1030(a)(2) sentencing by access type

<i>Access Type</i>	<i>Prison</i>	<i>Median Months</i>	<i>Restitution</i>	<i>N</i>
Exceeded	18.1%	7	20.8%	221
Unauthorized	26.7%	12	51.6%	161
Unknown	12.8%	4	38.5%	39
Total	20.9%	9	34.2%	421

Note: “Prison” lists the percentage of sentences that included a term of imprisonment. “Median months” lists the median length of imprisonment for sentences that included any prison term. This table excludes 1030(a)(2) sentences with non-CFAA statutes of conviction (except 18 U.S.C. § 2 (principals), § 371 (conspiracy), § 3571 (sentence of fine), or petty offenses). Source: PACER data set.

²⁹⁰ The 422 (a)(2) sentences include 6 that appeared as co-statutes of conviction with other 1030(a) subsections.

As shown in Table 15, employee access in excess of authorization stands out for its lighter sentencing than other (a)(2) fact patterns. As discussed above, these cases involved employees who were authorized to use a computer system but used it for an unauthorized purpose, such as looking at a tax return without a valid reason or feeding confidential information to someone. In the vast majority of these cases (85%), the defendant received no prison time. Unauthorized access by former employees was also unlikely to lead to prison or jail time. In contrast, 45% of the 22 defendants other than employees who exceeded authorized access received jail or prison terms.

Table 15: 1030(a)(2) sentencing by fact pattern

<i>Fact Pattern</i>	<i>Prison</i>	<i>Median Months</i>	<i>Restitution</i>	<i>N</i>
Employee	17.0%	9	18.0%	206
Other use in excess of authorization	33.3%	6	60.0%	15
Snooping	30.0%	8	26.0%	50
Former employee	16.7%	12	72.2%	54
Other unauthorized access	33.3%	12	54.4%	57
Unknown access type	12.8%	4	38.5%	39
Total	20.9%	9	34.2%	421

Note: “Prison” lists the percentage of sentences that included a term of imprisonment. “Median months” lists the median length of imprisonment for sentences that included any prison term. This table excludes 1030(a)(2) sentences with non-CFAA statutes of conviction (except 1030 § 2 (principals), § 371 (conspiracy), § 3571 (sentence of fine), or petty offenses). Source: PACER data set.

4.3.2.5. Co-Statutes of Conviction with 1030(a)(2)

Table 16 lists the co-statutes of conviction with 1030(a)(2) that occur four times or more in the data set. One of the interesting things about Table 16 is the number of 1030(a)(2) convictions that appear alongside convictions for fraud. Fact patterns involving fraud could presumably be charged under 1030(a)(4), which applies specifically to unauthorized access in furtherance of fraud. So one might expect these statutes to appear with 1030(a)(4) convictions—and indeed, fraud statutes appear with 1030(a)(4) convictions 101 times, representing 42% of the 242 1030(a)(4) convictions and 89% of the 112 (a)(4) convictions with multiple statutes of conviction.²⁹¹ But fraud statutes also appear 89 times with 572 1030(a)(2) convictions (16%) and in 59% of the 150 1030(a)(2) convictions with multiple statutes.²⁹² Section 1030(a)(2) may not be a fraud statute, but it frequently appears together with fraud convictions.

²⁹¹ For the purposes of this analysis, I consider the following statutes to be fraud statutes: 18 U.S.C. §§ 1001, 1028, 1028A, 1029, 1341–1349, 1956, and 1519.

²⁹² These figures include sentences with convictions under multiple 1030 subsections, including four sentences with convictions under both (a)(2) and (a)(4).

Table 16: Co-statutes of conviction with 1030(a)(2)

<i>USC Section</i>	<i>Description</i>	<i>N</i>
18:2	Punishment as a principal	88
18:1028A	Identity theft or aggravated identity theft ²⁹³	60
18:371	Conspiracy	54
18:1344	Bank fraud	30
18:1029	Access-device fraud	23
18:1343	Wire fraud	12
18:2252	Child pornography	10
18:2511	Interception of communications (wiretap)	10
18:1001	False statements	8
18:1028	Identity theft	8
18:1349	Attempted fraud	8
18:875	Extortionate interstate communications	8
18:201	Bribery of public officials and witnesses	6
18:2261	Interstate domestic violence	6
18:1346	Definition of “scheme or artifice to defraud”	5
18:1512	Witness tampering	5
18:1341	Mail fraud	4
18:1503	Influencing or injuring officer or juror	4
26:7213	Unauthorized distribution of tax information	4
Others		73
Total		426

Note: N counts the number of times a given statute appeared with 1030(a)(2) as a statute of conviction.

Source: USSC data set supplemented by PACER data.

4.3.2.6. “Special Skills” and “Sophisticated Means”

The data set lets us look at a related sentencing issue for which the data has previously been “sparse”:²⁹⁴ whether the sentencing enhancements for “sophisticated means” and “special skills” are commonly applied in computer crime cases. The use of these enhancements in computer crime cases is controversial; Orin Kerr, for example, has written that these enhancements are “hard to justify” but appear to be widely used in “run-of-the-mill CFAA cases.”²⁹⁵

The “sophisticated means” enhancement adds two levels to the offense level of economic crimes when the crime involved “especially complex or especially intricate offense conduct pertaining to the execution or concealment of an offense.”²⁹⁶ The “special skills” enhancement adds two offense levels when “the defendant abused a position of public or private trust, or used a special skill, in a manner that significantly facilitated the commission or concealment of the offense.”²⁹⁷ A “special skill” is a skill “not possessed by members of the general public and

²⁹³ Because of the way the USSC data set encodes statutes, it is not possible to distinguish convictions under the aggravated identity theft statute, 18 U.S.C. § 1028A, from those under subsection (a) of the identity theft statute, 18 U.S.C. § 1028(a).

²⁹⁴ Kerr, *Trespass, Not Fraud*, *supra* note 208, at 1562.

²⁹⁵ *Id.* at 1562–63.

²⁹⁶ U.S. SENTENCING GUIDELINES MANUAL §§ 2B1.1(b)(10)(C), cmt. 9(B).

²⁹⁷ *Id.* § 3B1.3.

usually requiring substantial education, training or licensing”; examples include “pilots, lawyers, doctors, accountants, chemists, and demolition experts.”²⁹⁸ The “special skills” enhancement is not supposed to be used when the abuse of trust or skill is included in the base offense level.²⁹⁹

Although the special-skills and sophisticated-means enhancements are by no means rare in CFAA sentences, their application does not appear to be common, either. According to the data set, the “special skills” enhancement was applied in 26% of all CFAA sentences (296/1,142) and 24% of sentences in which the CFAA was the only charge (203/829). The sophisticated-means enhancement was applied in 19% of all CFAA cases (199/1,054) and 14% of cases in which the CFAA was the only charge (109/793).³⁰⁰

Considering only (a)(2) sentences, the special-skills enhancement was applied more often in exceeding-authorization cases (37%) than in unauthorized-access cases (17%) (Table 17).

Table 17: Application of the “Special Skill” and “Sophisticated Means” enhancements by access type for (a)(2) sentences

<i>Access Type</i>	<i>Special Skill</i>	<i>N</i>	<i>Soph. Means</i>	<i>N</i>
Exceeded	37%	250	4%	229
Unauthorized	18%	212	21%	197
Unknown	20%	41	0%	34
Total	28%	503	11%	460

Note: Totals for *N* differ because the USSC data set did not have information for all sentences on whether the special-skill or sophisticated-means enhancements were applied. Source: Combined USSC and PACER data set.

It is a little surprising that the special-skills enhancement is used more often in exceeding-authorized-access cases than in unauthorized-access cases. As discussed in Section 4.3.2.1, the latter include the typical “hacking” cases that are more likely to involve “cleverness” by the offender in bypassing controls or escalating privileges. An offender who exceeded authorized access already had access to the information (but used that information for an improper purpose). I can only speculate as to the reason for this result, but it may be because of the “abuse of trust” aspect of the special-skills enhancement (i.e., the full enhancement is for using a special skill *or* abusing a position of trust). The special skills enhancement was applied in 40% of (a)(2) cases with the employee fact pattern (Table 18), suggesting that abuse of position might have been a factor. The enhancement was applied only 18% of the time in non-employee-EAA cases ($p < 0.001$, χ^2). But it could also be that the special skills are “baked into” unauthorized access fact patterns, while exceeding access has more room to use a special skill or abuse a position of trust beyond what’s already part of the crime.

The “sophisticated means” enhancement, on the other hand, was applied much more often in unauthorized access cases than in exceeded-authorization cases. It was applied in 31%

²⁹⁸ *Id.* § 3B1.3, cmt. 4.

²⁹⁹ *Id.*

³⁰⁰ The total number of cases differ between the special-skills and sophisticated-means totals because not all records in the USSC dataset included information on whether these enhancements were applied.

of “snooping” sentences and 23% of other types of UA sentences, but in only 9 (4%) of 229 exceeding-authorized-access cases.³⁰¹ This makes some sense: if someone is already authorized to use a computer system, exceeding that access by using information for an unauthorized purpose does not require any particular sophistication.

Table 18: Application of the “Special Skill” and “Sophisticated Means” enhancements by fact pattern for (a)(2) sentences

<i>Fact Pattern</i>	<i>Special Skill</i>	<i>N</i>	<i>Soph. Means</i>	<i>N</i>
Employee	40%	231	4%	212
Other use in excess of authorization	15%	20	6%	17
Snooping	16%	75	31%	67
Former employee	19%	57	5%	55
Other unauthorized access	18%	79	23%	74
Unknown access type	20%	41	0%	34
Total		503		459

Note: Totals for *N* differ because the USSC data set did not have information for all sentences on whether the special-skill or sophisticated-means enhancements were applied. Source: Combined USSC and PACER data set.

4.3.3. What do Federal Trespass Crimes Look Like?

There is no general federal trespass statute. Instead, a few laws and regulations prohibit trespass and similar activities in specific places. These places include public lands (e.g., national

³⁰¹ See Factual Statement, *United States v. Allison*, No. 07-cr-00016 (E.D. Tex. June 28, 2007) (accessing and changing credit files w/o authorization); Plea Agreement, *United States v. Alavi*, No. 07-cr-00429 (D. Ariz. Dec. 16, 2008) (copying nuclear simulator software before leaving employer and taking it to Israel); Plea Agreement, *United States v. Williams*, No. 09-cr-00298 (M.D. Fla. June 26, 2009) (using Western Union wire transfers to receive proceeds of credit card fraud in violation of WU terms of service); Plea Agreement, *United States v. Plom*, No. 11-cr-05173 (W.D. Wash. Mar. 29, 2011) (government employee providing wage & employment data to a third party in exchange for payment); Indictment, *United States v. Merrick*, No. 10-cr-00572 (E.D. Pa. Sept. 2, 2010) (bank employee providing account information to others who used it for bank fraud); Stipulated Factual Basis, *United States v. Woodruff*, No. 12-cr-00198 (S.D. Ind. Jan. 16, 2014) (as part of mechanic's lien fraud scheme, asking DMV employee to access motor vehicle info w/o proper purpose); Plea Agreement, *United States v. Lorash*, No. 12-cr-01008 (N.D. Ill. Dec. 18, 2013) (insurance agent downloading proprietary files with intent to sell them to competitors); Indictment, *United States v. Guevara et al*, No. 14-cr-00649 (S.D. Tex. Apr. 29, 2014) (Texas Workforce Commission employee entering false information so co-conspirator would receive fraudulent UI benefits); Plea Agreement, *United States v. Hall*, No. 15-cr-00314 (N.D. Ala. Feb. 8, 2016) (IRS agent accessing tax return data and using it to file fraudulent 1040 returns). *Allison*, *Williams*, *Lorash*, and *Guevara* all received probation, so it is unlikely that the enhancement made any difference in the final sentence. *Hall* was an atypical IRS agent case: instead of just looking at tax forms, *Hall* “used fraudulently obtained dates of birth, social security numbers, and names . . . to file hundreds of fraudulent IRS 1040 individual income tax forms” to obtain over \$430,000 in fraudulent refunds. Plea Agreement at 5–6, *Hall*, No. 15-cr-00314.

parks and forests),³⁰² federal buildings and property,³⁰³ Indian lands,³⁰⁴ critical infrastructure facilities,³⁰⁵ and ships, aircraft, and airports.³⁰⁶ Although some of the statutes prohibit specific acts, many of the statutes cover assorted forms of misconduct on public land. Trespass is sometimes charged in conjunction with these other offenses.

The penalties for federal trespass vary greatly, from a low of a \$1000 (with no prison sentence) for trespassing on a nuclear facility with a handgun or explosive to life in prison for intentionally causing death by stowing away on a ship or aircraft. The only federal trespass crimes that are felonies are trespassing on airports or aircraft and stowing away on ships or aircraft.³⁰⁷ Several of the trespass crimes are Class B or C misdemeanors—also called “petty offenses.”³⁰⁸ These do not appear in the USSC data set because it includes only convictions for felonies and Class A misdemeanors,³⁰⁹ which are also the only crimes to which the sentencing guidelines apply.³¹⁰ A spot check confirmed that some petty offenses appear in PACER court files, but these tend to have much less information about the facts of each case than do Class A misdemeanor and felony cases.

Public Lands

Public lands include national parks, national forests, wildlife refuges, and similar lands such as national cemeteries.

The statute covering trespass in national parks, 43 U.S.C. § 1733, encompasses a wide range of conduct. Section 1733 does not prohibit any conduct directly. Instead, it requires the Secretary of the Interior to “issue regulations . . . with respect to the management, use, and

³⁰² See 16 U.S.C. § 668dd (wildlife refuges); 50 C.F.R. § 26.21; 18 U.S.C. § 1857 (driving livestock onto public lands); 18 U.S.C. § 1863 (national forests); 38 U.S.C. § 2413 (national cemeteries); 43 U.S.C. § 1733 (national parks).

³⁰³ See 18 U.S.C. § 1036 (entering federal property by false pretenses); 18 U.S.C. § 1382 (military lands); 18 U.S.C. § 1752 (restricted buildings or grounds); 18 U.S.C. § 1793 (federal prison lands).

³⁰⁴ See 18 U.S.C. § 1165 (hunting or fishing on Indian land); 25 U.S.C. § 3106 (forests), § 3713 (agricultural land); 25 C.F.R. § 11.411.

³⁰⁵ See 42 U.S.C. § 2278a (nuclear facilities); 42 U.S.C. § 7270b (strategic petroleum reserve).

³⁰⁶ See 18 U.S.C. § 2199 (ships and aircraft); 49 U.S.C. § 46314 (airports and aircraft).

³⁰⁷ The stowaway statute, 18 U.S.C. § 2199, prohibits boarding or remaining aboard an aircraft or vessel “with intent to obtain transportation.” By contrast, 49 U.S.C. § 46314 prohibits knowingly and willfully entering an aircraft or airport area “in violation of security requirements.”

³⁰⁸ 18 U.S.C. § 19. Class B misdemeanors are offenses for which the maximum sentence is “six months or less but more than thirty days.” Class C misdemeanors are offenses for which the maximum sentence is “thirty days or less but more than five days.” 18 U.S.C. § 3559 (classifying offenses).

³⁰⁹ LOU REEDT, COURTNEY SEMISCH & KEVIN BLACKWELL, U.S. SENTENCING COMM’N, EFFECTIVE USE OF FEDERAL SENTENCING DATA 5 (Nov. 2013). A Class A misdemeanor is an offense for which the maximum punishment is “one year or less but more than six months.” 18 U.S.C. § 3559. Offenses with a maximum sentence of more than one year are classified as felonies (which are also divided into Classes A to E, but the distinctions between those levels is not important here).

³¹⁰ U.S. SENTENCING GUIDELINES MANUAL § 1B1.9 (stating that “[t]he sentencing guidelines do not apply to any count of conviction that is a Class B or C misdemeanor or an infraction”).

protection of the public lands.”³¹¹ Thus, the behavior that is specifically prohibited is set out in the Code of Federal Regulations.³¹² Cases charged under the statute include a fairly broad range of things people are not supposed to do on public lands, including illegally setting fires,³¹³ destroying trees or plants,³¹⁴ operating motorcycles without valid registrations,³¹⁵ hunting without a permit,³¹⁶ being in illegal possession of firearms,³¹⁷ littering,³¹⁸ and camping for longer than is allowed.³¹⁹

The wildlife refuge statute, 16 U.S.C. § 668dd, is similarly broad. Although the statute specifically states that no person shall “enter, use, or otherwise occupy any [wildlife refuge] area for any purpose” without authorization,³²⁰ it also prohibits taking or damaging property, plants, or animals on wildlife refuges.³²¹ It further authorizes the Secretary of the Interior to promulgate regulations prohibiting other conduct. These prohibitions include, for example, gambling,³²² indecent exposure,³²³ violation of state and local vehicle laws,³²⁴ littering,³²⁵ and possession of controlled substances,³²⁶ to name a few. None of the convictions under 16 U.S.C. § 668dd since 2005 were sentenced under the guideline for trespass.³²⁷

Two national parks, Crater Lake National Park³²⁸ and Wind Cave National Park,³²⁹ contain provisions in their establishing statutes that prohibit trespass. Each provision carries a one-year maximum sentence. Neither statute appears in the dataset. Another statute that appears not to have been used is 18 U.S.C. § 1857, which provides for up to one year imprisonment for

³¹¹ 43 U.S.C. § 1733(a).

³¹² *See, e.g.*, 43 C.F.R. § 9262.1, 9264.1.

³¹³ *See, e.g.*, Violation Notice, United States v. Pretzer, No. 1:13-mj-00130 (D. Colo. June 30, 2013).

³¹⁴ *See, e.g.*, Docket, United States v. Wallace, No. 1:14-mj-00102 (D. Colo. Apr. 18, 2014).

³¹⁵ *See, e.g.*, Misdemeanor Information, United States v. Kester, No. 2:09-cr-00113 (D. Utah Mar. 2, 2009).

³¹⁶ *See, e.g.*, United States v. Bishop, No. 1:06-mj-00110 (D. Colo. Mar. 7, 2006).

³¹⁷ *See, e.g.*, Criminal Complaint, United States v. Herrera-Sanchez, No. 6:07-cr-60055 (D. Ore. Apr. 1, 2007).

³¹⁸ *See, e.g.*, Criminal Complaint, United States v. Pinotes, No. 1:08-mj-00108 (D. Colo. Mar. 14, 2008).

³¹⁹ *See, e.g., id.*; Criminal Complaint, United States v. Owen, No. 1:10-mj-00130 (D. Colo. July 12, 2010) (alleging that the defendant was living on land managed by the Bureau of Land Management).

³²⁰ 16 U.S.C. 668dd(c).

³²¹ *Id.*

³²² 50 C.F.R. § 27.85.

³²³ 50 C.F.R. § 27.83. *See also, e.g.*, Information, United States v. Underwood, No. 5:09-cr-00382 (N.D. Ala. Sept. 2, 2009); Information, United States v. Young, No. 5:09-cr-00383 (N.D. Ala. Sept. 2, 2009).

³²⁴ 50 C.F.R. § 27.31.

³²⁵ 50 C.F.R. § 27.94.

³²⁶ 50 C.F.R. § 27.82. *See also, e.g.*, Violation Notice, United States v. Stone, No. 3:09-po-00002 (D. Alaska June 12, 2009).

³²⁷ Eighty of those convictions were in two districts: the Northern District of Alabama and the Western District of Oklahoma.

³²⁸ 16 U.S.C. § 123 (establishing a penalty of up to \$500 fine or one year imprisonment for anyone who “establish[es] any settlement or residence” or “engage[s] in any lumbering, or other enterprise” within the park).

³²⁹ 16 U.S.C. § 146 (establishing a penalty of up to \$1000 or one year imprisonment for unlawfully intruding into the park).

anyone who knowingly permits “cattle, horses, hogs, or other livestock” to enter into fenced public lands where they may cause property damage on those lands.³³⁰

I found no cases for trespassing in a national forest, 18 U.S.C. § 1863, or for demonstrations within two hours of a funeral at a national cemetery, 38 U.S.C. § 2413.

Federal Property

The “trespass” statute covering military bases, 18 U.S.C. § 1382, tends to accompany other charges. It provides for a fine or imprisonment of up to six months for anyone “who goes upon any military, naval, or Coast Guard reservation, post, fort, arsenal, yard, station, or installation, for any purpose prohibited by law or lawful regulation.” The language “for any purpose prohibited by law or lawful regulation” makes the crime somewhat similar to a burglary offense—the crime is entering onto military land with the intent of committing any crime therein. Thus, section 1382 is often charged along with other offenses such as theft.³³¹

The statute for trespassing on “any restricted building or grounds” applies to attempts to enter the White House, the Vice President’s residence, or any place where someone protected by the Secret Service is or will be visiting.³³² Thus, the handful of convictions under this statute involved people climbing the White House fence,³³³ trying to get to someone protected by the Secret Service,³³⁴ or entering a restricted area at a political convention.³³⁵

18 U.S.C. § 1036 prohibits entering into any “real property, vessel, or aircraft of the United States.” Fact patterns under this section appear to vary by jurisdiction. In Utah, it seems to be another statute used for people carrying weapons into airports, as a plea down from 49 U.S.C. § 46505, which has at least a 10 year maximum sentence.³³⁶ In the Southern District of Georgia, section 1036 has been used against people trying to enter military bases with invalid identification cards.³³⁷

18 U.S.C. § 1793 prohibits trespassing on Bureau of Prisons land. I was unable to find any cases sentenced under this statute.

³³⁰ See 18 U.S.C. § 1857. Although the USSC data set shows three sentences under 18 U.S.C. § 1857 since 2005, the controlling sentencing guideline in all three was § 2S1.1, which applies to money laundering offenses. I think it highly likely that these were simply data entry errors for racketeering crimes under 18 U.S.C. § 1957.

³³¹ See, e.g., Information, United States v. Reas, No. 3:11-mj-00324 (W.D. Ky. Oct. 20, 2011).

³³² 18 U.S.C. § 1752.

³³³ See, e.g., Statement of the Offense, United States v. Caputo, No. 15-cr-175 (D.D.C. Sept. 12, 2016); Statement of the Offense, United States v. Gozalez, No. 14-cr-200 (D.D.C. Mar. 13, 2015).

³³⁴ See, e.g., Complaint, United States v. Ernst, No. 14-cr-286 (D. Nev. Apr. 13, 2014) (entering area of a hotel cordoned off for a visit by Hillary Clinton).

³³⁵ See Complaint, United States v. Martin, No. 16-cr-465 (E.D. Pa. July 28, 2016) (entering restricted grounds of the Democratic National Convention by cutting through a fence).

³³⁶ 49 U.S.C. § 46505(b). In probably the most memorable case, the defendant had used his twin brother’s airport ID to work in the airport under his twin’s name. Plea Statement, United States v. Nedelcu, No. 05-cr-00330 (D. Utah Sept. 8, 2005),

³³⁷ See, e.g., Information, United States v. Franks, No. 18-cr-00277 (S.D. Ga. Dec. 7, 2018) (showing a photo of an expired identification card to gain access to a military base).

The maximum sentence for trespassing on federal property is generally six months, although the maximum can increase to ten years if aggravating factors are present.³³⁸

Indian Lands

Three federal statutes prohibit trespassing on Indian lands. Only one of them is a criminal statute, however: 18 U.S.C. § 1165 provides for up to a 90 day sentence for hunting, trapping, or fishing on Indian lands. The two other statutes provide for civil monetary penalties for trespassing in Indian forest lands, 25 U.S.C. § 3106, and Indian agricultural lands, 25 U.S.C. § 3713.³³⁹

Ships, Aircraft, and Airports

The stowaway statute, 18 U.S.C. § 2199, prohibits boarding or remaining aboard an aircraft or vessel “with intent to obtain transportation.” All the cases I reviewed for this statute involved foreign nationals stowing away on vessels bound for the United States.³⁴⁰ Although the statute allows for a sentence of up to five years for offenses that do not result in injury, the sentence in each case I saw was time served and delivery to the United States Marshals Service (presumably for deportation).

The cases under 49 U.S.C. § 46314 also had a common character. That statute prohibits knowingly and willfully entering an aircraft or airport area “in violation of security requirements.”³⁴¹ The cases I reviewed involved people going through airport security with guns.

Critical Infrastructure

These statutes prohibit trespassing on nuclear facilities, 42 U.S.C. § 2278a, and the strategic petroleum reserve, 42 U.S.C. 7270b. Although eleven sentences have been imposed for the former since 2005,³⁴² I was unable to find any case documents for trespassing on nuclear facilities or federal prison lands.

³³⁸ 18 U.S.C. § 1752(b) provides for a ten year maximum sentence if the trespasser carries a dangerous weapon or the offense results in significant bodily injury. 18 U.S.C. § 1036 carries a ten-year maximum sentence if the trespass was committed “with the intent to commit a felony.”

³³⁹ The civil penalties are established by Department of the Interior regulations. *See* 25 C.F.R. § 163.29 (forests); 25 C.F.R. §§ 166.800–819 (grazing).

³⁴⁰ *See, e.g.,* Factual Proffer Supporting Change of Plea, *United States v. Ramirez-Torres*, No. 0:11-cr-60193 (S.D. Fla. 2011); Factual Basis, *United States v. Bonilla*, No. 2:09-cr-00045 (E.D. La. May 13, 2009); Factual Basis, *United States v. Cardenas*, No. 2:09-00326 (E.D. La. Dec. 16, 2009).

³⁴¹ 49 U.S.C. § 46314.

³⁴² All but one of the 11 sentences for trespassing on nuclear facilities were in one district in one year: the Eastern District of Tennessee in 2011. A year later, a group of protesters, including an 82-year-old nun, were charged with several crimes, including trespassing under 42 U.S.C. § 2278a, after they “cut through four layers of fences” and spray-painted antiwar slogans on a storage facility for enriched uranium. *United States v. Walli*, 785

4.3.4. What do Federal Burglary Crimes Look Like?

As with trespass, there is no generally applicable federal burglary statute. But federal burglary cases are more prevalent than trespass crimes at the misdemeanor and felony level. They are also fit the typical meaning of the crime better—while crimes involving stowaways and violating airport security requirements stretch the definition of “trespassing,” the federal burglary crimes fall more solidly within the typical definition of the crime. The traditional definition of burglary, at common law, is breaking and entering the dwelling of another at night with the intent to commit a felony therein.³⁴³ Modern burglary statutes have expanded the definition to include the entry of a building with the intent to commit a crime,³⁴⁴ abandoning requirements that limited common-law burglary to dwellings,³⁴⁵ daylight,³⁴⁶ and felonies.³⁴⁷ It is not necessary that a burglar actually commit a crime—only the intent to commit a crime is needed.

Contrast the definition of burglary with robbery, which at common law is (reduced to its essentials) a theft by force or threat of force.³⁴⁸ Robbery, therefore, is an inapt analogue to CFAA (a)(2) crimes—the threat of physical violence makes robbery a more serious crime than unauthorized access to a computer. Burglary, however, may be perceived as being very much like a CFAA (a)(2) case.

There are four main federal burglary statutes, covering burglary of banks, post offices, carriers, and burglary of controlled substances. Section 2113(a) prohibits entering a “bank, credit union, or any savings and loan association” with the intent to commit a felony or larceny.³⁴⁹ Section 2115 sets criminal penalties for anyone who “forcibly breaks into or attempts to break into any post office . . . with intent to commit . . . any larceny or other depredation.”³⁵⁰ Cases sentenced under that statute involve people breaking into post offices and stealing boxes of checks,³⁵¹ drugs,³⁵² and, of course, mail.³⁵³ Section 2117 prohibits breaking into *or* entering “any railroad car, vessel, aircraft, motortruck, wagon or other vehicle or of any pipeline system, containing interstate or foreign shipments . . . with intent in either case to commit larceny

F.3d 1080, 1083 (6th Cir. 2015). Prosecutors eventually dropped the trespassing charges, which may be why that case did not come up in a PACER criminal case search for 42:2278A.M.

³⁴³ E.g., *Taylor v. United States*, 495 U.S. 575, 580 n.3 (1990), *quoting* W. LAFAVE & A. SCOTT, SUBSTANTIVE CRIMINAL LAW § 8.13, p. 464 (1986) (“Burglary was defined by the common law to be the breaking and entering of the dwelling house of another in the nighttime with the intent to commit a felony.”).

³⁴⁴ See Helen A. Anderson, *From the Thief in the Night to the Guest Who Stayed Too Long: The Evolution of Burglary in the Shadow Of the Common Law*, 45 IND. L. REV. 629 (2012) (surveying state burglary laws).

³⁴⁵ *Id.* at 647–49.

³⁴⁶ *Id.* at 642–44.

³⁴⁷ *Id.* at 651–52.

³⁴⁸ E.g., *Carter v. United States*, 530 U.S. 255, 275–80 (2000) (Ginsberg, J., dissenting).

³⁴⁹ 18 U.S.C. § 2113(a).

³⁵⁰ 18 U.S.C. § 2115.

³⁵¹ See Indictment, *United States v. Neyland*, Np. 3:14-cr-00287 (N.D. Tex. July 22, 2014).

³⁵² See Factual Résumé, *United States v. West*, No. 6:07-cr-00012 (N.D. Tex. May 17, 2007).

³⁵³ See Factual Resume, *United States v. Harris*, No. 2:09-cr-00169 (S.D. Ala. Oct. 20, 2009).

therein.”³⁵⁴ The cases I found all involved people breaking into railroad cars and stealing cargo.³⁵⁵ And the cases under section 2118(b), which establishes penalties for anyone who “without authority, enters or attempts to enter, or remains in, the business premises or property of a person registered with the Drug Enforcement Administration . . . with the intent to steal any material or compound containing any quantity of a controlled substance,” unsurprisingly involve people breaking into pharmacies to steal drugs.³⁵⁶

The federal burglary crimes are all felonies with maximum sentences ranging from 5 years to 20 years for a first offense. These are as high or higher than the maximum penalties for first-offense CFAA (a)(2) crimes. The sentencing guidelines may also call for harsher sentences depending on the amount of loss. The base offense level is 17 (corresponding to a sentence of about 27 months for a first-time offender) for burglarizing a residence and 12 (corresponding to a sentence of about 13 months for a first-time offender) for burglarizing a non-residence.³⁵⁷ Enhancements based on the amount of loss progress less sharply for burglary than for fraud, however. For example, burglary resulting in \$200,000 in loss would receive a three-point enhancement for a total offense level of 15 before any other applicable adjustments.³⁵⁸ A computer intrusion resulting in the same loss would receive a ten-point enhancement for a total offense level of 16 before other adjustments.³⁵⁹

The burglary sentencing guidelines have also been used under various other statutes, including those that apply to gun offenses³⁶⁰ or broadly cover any crimes committed in Indian lands.³⁶¹ Section 2B2.1 was also applied for statutes that describe roles,³⁶² apply state laws to crimes on federal lands,³⁶³ or cover crimes in maritime and special territorial jurisdiction.³⁶⁴

4.3.5. What do Federal Fraud Crimes Look Like?

Unlike trespass, burglary, and eavesdropping, there are many federal fraud statutes and sentences. The statutory index to the sentencing guidelines lists 295 statutes that may be sentenced according to section 2B1.1, which covers fraud, larceny, theft, and other economic crimes. Section 2B1.1 is the third-most sentenced section in the USSC data set, accounting for

³⁵⁴ 18 U.S.C. § 2117.

³⁵⁵ See, e.g., Plea Agreement, United States v. Dushaj, No. 2:12-cr-20764 (E.D. Mich. Feb. 15, 2013) (car parts); Plea Agreement, United States v. Cross, No. 2:12-cr-20831 (E.D. Mich. Mar. 28, 2013) (whiskey).

³⁵⁶ See, e.g., Information, United States v. Bendyna, No. 2:11-cr-00317 (E.D. Pa. June 2, 2011).

³⁵⁷ U.S. SENTENCING GUIDELINES MANUAL § 2B2.1.

³⁵⁸ *Id.*

³⁵⁹ *Id.* § 2B1.1.

³⁶⁰ 18 U.S.C. §§ 922, 924.

³⁶¹ 18 U.S.C. §§ 1151, 1152, 1153.

³⁶² 18 U.S.C. §§ 2, 371.

³⁶³ 18 U.S.C. § 13.

³⁶⁴ 18 U.S.C. § 7.

nearly 12% of all convictions.³⁶⁵ Some of the most commonly sentenced fraud crimes include mail fraud,³⁶⁶ wire fraud,³⁶⁷ bank fraud,³⁶⁸ health care fraud,³⁶⁹ embezzlement,³⁷⁰ identity theft,³⁷¹ access device fraud,³⁷² Social Security fraud,³⁷³ false claims to government funds,³⁷⁴ and counterfeiting a security.³⁷⁵ Maximum sentences for first-time offenders range from one year (for embezzlement of public money or by a bank employee of \$1,000 or less)³⁷⁶ to life (for healthcare fraud that results in death),³⁷⁷ but maximum sentences of 5, 10, 20, or 30 years are typical.³⁷⁸ Notably, two crimes that might appear similar to CFAA crimes—mail fraud and wire fraud—have maximum sentences of 20 years for a first offense, which can increase to 30 years with aggravating factors.³⁷⁹ Another potentially relevant fraud crime, identity fraud, has a maximum sentences that starts at 1 year but can increase to 5, 15, 20, or 30 years based on the presence of aggravating facts.³⁸⁰

³⁶⁵ The two most-sentenced sections of the Guidelines are § 2D1.1 (33% of all convictions), covering drug trafficking, and § 2L1.2 (23% of all convictions), which is used for crimes involving “unlawfully entering or remaining in the United States.”

³⁶⁶ 18 U.S.C. § 1341 (mail fraud); 18 U.S.C. § 1708 (mail theft); 18 U.S.C. § 1709 (mail theft by a postal employee).

³⁶⁷ 18 U.S.C. § 1343.

³⁶⁸ 18 U.S.C. § 1344.

³⁶⁹ 18 U.S.C. § 1347.

³⁷⁰ 18 U.S.C. § 641 (embezzlement of public money); 18 U.S.C. § 656 (embezzlement by bank employee).

³⁷¹ 18 U.S.C. § 1028 (identity fraud); 18 U.S.C. § 1028A (aggravated identity theft).

³⁷² 18 U.S.C. § 1029. An “access device” is essentially anything that allows access to a financial or credit account, such as a credit card, account number, PIN, or “any other means of access that can be used . . . to obtain money, good, services, or any other thing of value.” 18 U.S.C. § 1029(e)(1).

³⁷³ 42 U.S.C. § 408.

³⁷⁴ 18 U.S.C. § 287 (false claims); 18 U.S.C. § 286 (false claims conspiracy).

³⁷⁵ 18 U.S.C. § 513.

³⁷⁶ 18 U.S.C. §§ 641, 656.

³⁷⁷ 18 U.S.C. § 1347 (“[I]f the violation results in death, such person shall be fined under this title, or imprisoned for any term of years or for life, or both.”).

³⁷⁸ Fraud statutes with 5-year maximum sentences include 18 U.S.C. § 1001 (false statements), 18 U.S.C. § 1029 (access device fraud), 18 U.S.C. § 1708 (mail theft), 18 U.S.C. § 1709 (mail theft by a postal employee), and 42 U.S.C. § 408 (social security fraud). Statutes with 10-year maximum sentences include 18 U.S.C. § 286 (false claims conspiracy), 18 U.S.C. § 513 (counterfeiting a security), 18 U.S.C. § 641 (embezzlement of more than \$1,000 in public money), 18 U.S.C. § 656 (embezzlement of more than \$1,000 by a bank employee), 18 U.S.C. § 666 (theft or bribery related to federally-funded programs), 18 U.S.C. § 1711 (misappropriation of postal funds), and 18 U.S.C. § 2314 (transportation of stolen goods). Statutes with 20-year maximum sentences include 18 U.S.C. § 1341 (mail fraud), 18 U.S.C. § 1343 (wire fraud), and 18 U.S.C. § 1347 (health care fraud).

³⁷⁹ The 30-year maximum sentence applies “[i]f the violation occurs in relation to, or involving any benefit authorized, transported, transmitted, transferred, disbursed, or paid in connection with, a presidentially declared major disaster or emergency . . . or affects a financial institution.” 18 U.S.C. §§ 1341, 1343.

³⁸⁰ Maximum sentences for identity fraud can increase to 5 or 15 years for producing (as opposed to merely using) fraudulent identification documents, 18 U.S.C. § 1028(b)(1)–(2), 20 years if committed in connection with a drug crime or violent crime, § 1028(b)(3), or 30 years if “committed to facilitate an act of domestic terrorism,” § 1028(b)(4). The sentencing for *aggravated* identity theft is something of a special case: aggravated identity theft must be in connection with another felony, and adds up to 2 years or 5 years to the sentence, which is not supposed

Because there have been so many fraud convictions since 2005 (133,012 sentences were imposed for which section 2B1.1 set the highest sentence, according to the USSC data set), a comprehensive review of court documents for these cases was not feasible. I therefore used court records³⁸¹ to review a small, non-random sample of fraud cases.

The range of fraud statutes is varied, but many of the statutes are fairly clear, if not necessarily specific, about the conduct they prohibit. For example, mail fraud covers basically any fraud committed using the mail.³⁸² Wire fraud is fraud that uses wire or radio communication.³⁸³ Bank fraud is fraud against a financial institution.³⁸⁴ And so on.

Although fraud statutes may be simple, the fact patterns they cover can be complex. For example, one of the simpler mail fraud cases I found in my review of cases was a controller and general manager who, over a ten-year period, used her check-signing authority to write approximately \$1.5 million in checks drawn on her employer's account. The checks were ostensibly for paying vendors but actually went to pay for her personal use.³⁸⁵ In another case, the defendant was charged with selling "time deposit certificates," which the defendant claimed were interest-bearing debt instruments issued by legitimate financial institutions. She used the money she collected selling these "certificates" to pay the business expenses of companies she owned, for personal expenses, and to make payments to earlier investors.³⁸⁶

Department of Justice policy also affects the cases that appear as fraud convictions. According to the U.S. Attorneys' Manual, mail and wire fraud prosecutions "ordinarily should not be undertaken if the scheme employed consists of some isolated transactions between individuals, involving minor loss to the victims," but a "scheme which in its nature is directed to defrauding a class of persons, or the general public, with a substantial pattern of conduct" should be given "serious consideration" for prosecution.³⁸⁷ This emphasis on prosecuting large-scale mail fraud skews federal convictions toward crimes with greater punishments.

to be concurrent (although judges have discretion to make the sentence concurrent; *see* 18 U.S.C. § 1028A(b)(4)). The higher maximum is for identity theft committed in connection with terrorism. *See* 18 U.S.C. § 1028A.

³⁸¹ Specifically, I used Bloomberg Law's docket search.

³⁸² *See* 18 U.S.C. § 1341 (establishing a maximum 20-year sentence for anyone who, as part of "any scheme or artifice to defraud," sends or receives mail via the US Postal Service or private interstate carrier).

³⁸³ *See* 18 U.S.C. § 1343 (covering "any scheme or artifice to defraud . . . by means of wire, radio, or television communication in interstate or foreign commerce").

³⁸⁴ *See* 18 U.S.C. § 1344 (establishing a 30-year maximum sentence for "defraud[ing] a financing institution").

³⁸⁵ *See* Plea Agreement, *United States v. Dyer*, No. 5:17-cr-00170, at 9 (Aug. 16, 2017).

³⁸⁶ *See* Information, *United States v. Campano*, No. 8:13-cr-00065, at 2–3 (May 13, 2013).

³⁸⁷ DEP'T OF JUSTICE, U.S. ATTORNEY'S MANUAL 9-43.100 (1997), <https://www.justice.gov/usam/usam-9-43000-mail-fraud-and-wire-fraud>.

4.3.6. CFAA Compared to Burglary, Trespass, and Fraud

4.3.6.1. Treatment Under the Sentencing Guidelines

Table 19 summarizes the sentencing guideline factors for fraud, burglary, and trespass crimes. As discussed more fully in Section 3.2.1.2, most CFAA offenses, including (a)(2) offenses, fall under section 2B1.1 of the guidelines, which covers theft, fraud, and similar economic crimes.³⁸⁸ Section 2B1.1 has a base offense level of 6, which can increase based on the amount of loss (from +2, for a loss of more than \$6500, to +30, for a loss of more than \$550 million), whether the crime had 10 or more victims (+2), whether the offense involved an intent to obtain personal information or involved unauthorized public release of personal information (+2), and whether the offense involved a government computer or a system used to maintain critical infrastructure (+2), among many other possible increases.

Trespass is covered by section 2B2.3, with a base offense level of 4. The offense level increases by 2 if the trespass was at a secure government facility, a nuclear energy facility, on “a vessel or aircraft of the United States,” in a secure area of an airport or seaport, at a national cemetery, at any restricted building or grounds, or on a government computer or a computer system used to maintain critical infrastructure. If the trespass is on a computer, the loss enhancement table from 2B1.1 also applies.

Burglary is covered by section 2B2.1. It has a base offense level of 12, or 17 if the burglary was of a residence. The burglary guideline increases the offense level based on the amount of loss, but it uses a different table than the one in 2B1.1 for fraud: it provides for an increase of +1 (for a loss of more than \$5000) to +8 (for a loss of more than \$9.5 million). The effect of the different loss tables is that burglary starts out with higher sentences than fraud until the amount of loss is over \$150,000, at which point fraud has the higher offense level.

Table 19: Sentencing Guideline Factors for Fraud, Burglary, and Trespass Crimes

<i>Crime Type</i>	<i>Guidelines Section</i>	<i>Base Offense Level</i>	<i>Relevant Offense Level Adjustments</i>
Fraud	2B1.1	6	Amount of loss: +2 to +30 CFAA & personal information: +2 CFAA and critical infrastructure: +2 to +6 Def. used “sophisticated means”: +2
Burglary	2B2.1	12 (17 if a residence)	Amount of loss: +1 to +8 Def. had a dangerous weapon: +2
Trespass	2B2.3	4	Residence or secure government installation: +2 Def. had a firearm: +2 CFAA ((a)(3)) & amount of loss: as per 2B1.1

³⁸⁸ U.S. SENTENCING GUIDELINES MANUAL appx. A (indexing statutes to sentencing guidelines sections). Section (a)(3), covering unauthorized access to government computers, is sentenced as a trespass crime. under 2B2.3 Threats to damage a computer are sentenced under 2B3.2, for extortion.

4.3.6.1. Demographics

Table 20 shows the demographics for CFAA crimes compared to non-CFAA fraud crimes, burglary, and trespass.³⁸⁹ To eliminate the effects of other statutes of conviction on sentences, this table and the others in this section compare only sentences in which there was a single count of conviction.³⁹⁰

Table 20: Demographics by crime type (single count of conviction)

<i>Crime Type</i>	<i>Male</i>	<i>White</i>	<i>Med. Age</i>
CFAA	80.8%	69.9%	37
Non-CFAA Fraud	64.3%	61.0%	40
Burglary	91.2%	38.3%	26
Trespass	82.3%	60.7%	33
All sentences	86.7%	70.8%	35

Note: “All sentences” includes sentences in the USSC data set with a single count of conviction from 2005 through 2018 (i.e., the crimes are not limited to CFAA, fraud, burglary, or trespass crimes). Source: Combined USSC and PACER data set.

The interesting thing about these demographics is not so much the CFAA demographics—which skew male and white in roughly the same percentages as for all crimes in the USSC database—but that some non-CFAA crime demographics differ quite a bit from the overall numbers. Non-CFAA fraud offenders, for example, are somewhat less likely to be male and have a slightly higher proportion who are non-white. The apparent low percentage of white burglary offenders is probably due to crimes on Indian land comprising a large number of these convictions.³⁹¹

4.3.6.2. Sentences Imposed

Table 21 summarizes sentencing across crime types (as above, the table compares only sentences in which there was a single count of conviction). The statistics show marked differences between how CFAA crimes and other crimes are sentenced. CFAA sentences differ not only from burglary and trespass crimes, but also from non-CFAA fraud crimes. CFAA sentences include prison time much less often than do any of the other comparison crimes, and when sentences include prison time, median CFAA sentences tend to be shorter than those for non-CFAA frauds or burglary—but longer than those for federal trespass crimes.

The sentencing differences between CFAA, burglary, and trespass crimes might be expected given the differences in sentencing guidelines and maximum sentences.³⁹² The

³⁸⁹ For the purposes of this section, I define a non-CFAA fraud sentence as any sentence in which section 2B1.1 was the controlling guideline (except for CFAA crimes), burglary crimes as those for which 2B2.1 was the controlling guideline, and trespass as those for which section 2B2.3 controlled.

³⁹⁰ Sentences with a single count of conviction represented 65.9% of the 1,209 CFAA cases, 71.4% of the 107,703 non-CFAA fraud cases, 80.7% of the 843 burglary cases, and 92.8% of the 83 trespass cases.

³⁹¹ See *supra* Section 4.3.4.

³⁹² See *supra* Section 3.2.1.

difference between CFAA sentences and non-CFAA fraud sentences is more surprising. Both CFAA and non-CFAA fraud crimes are sentenced under section 2B1.1 of the guidelines, after all. That the same section of the guidelines has been applied to CFAA and non-CFAA fraud with differing results suggests either that judges see these crimes as quite different despite their similar treatment in the guidelines, or that there is some other difference between them that is leading to one being sentenced more harshly than the other.

Table 21: Sentencing by crime type (single count of conviction)

<i>Crime Type</i>	<i>Prison</i>	<i>Median Months</i>	<i>N</i>
CFAA	35.0%	12	797
Non-CFAA Fraud	60.0%	15	76,875
Burglary	90.7%	21	680
Trespass	72.7%	3	77

Note: “Prison” lists the percentage of sentences that included a term of imprisonment. “Median months” lists the median length of imprisonment for sentences that included any prison term. This table includes only sentences for which there was a single count of conviction. Source: Combined USSC and PACER data set.

Of course, it is possible that the differences between crime types in Table 21 are simply the result of different distributions of crime severity within categories. For example, non-CFAA fraud crimes might have greater losses on average than CFAA crimes, or they might tend to be committed by repeat offenders more often. Two simple regression models allow for a rough analysis of this possibility, at least with respect to CFAA vs. non-CFAA fraud crimes: (1) a probit model for the binary variable indicating whether prison time was imposed, and (2) an OLS regression for the number of months imposed, if any.³⁹³ Both models use the offense level and offender’s criminal history score as controls. The offense level incorporates aspects of a crime’s severity through the adjustments set out in the guidelines, including the amount of loss, and thus acts as a proxy for many other variables that influence a crime’s severity.³⁹⁴ The criminal history category captures the offender’s prior criminal history and thus how “culpable” the offender is.³⁹⁵ The model also includes demographic controls for age, sex, race, and level of education.³⁹⁶ Some potential controls that are not subsumed by the offense level and criminal history category are also excluded, either in the interest of not overspecifying the model or because the variables have too many values. An example of the former is the amount of loss from the crime, which is not included in the model because its inclusion in the offense level calculation could result in a high degree of collinearity if both were used as controls. An example of the latter is jurisdiction:

³⁹³ Burglary and trespass crimes are omitted from the regression analysis because they are sentenced under different sections of the Guidelines than fraud, have different ranges of offense levels, and in the case of burglary crimes, uses a different loss table; including these would put too much strain on an already rudimentary model.

³⁹⁴ See U.S. SENTENCING GUIDELINES MANUAL § 2B1.1.

³⁹⁵ See U.S. SENTENCING GUIDELINES MANUAL ch. 4, pt. A, intro. cmt.

³⁹⁶ Specifically, the model uses the USSC data set variables AGE, MONSEX, NEWRACE (which simplifies the various race categories into White, Black, Hispanic, and other), and NEWEDUC (which collapses various education levels into less than high school, high school graduate, some college, and college graduate).

although it might be interesting to see whether CFAA sentences in some judicial districts are higher than in others, the large number of districts (97) and asymmetric distribution of cases among districts would slice the 797 CFAA cases (those with a single count of conviction) too finely.³⁹⁷

Table 22 shows the regression results. These results suggest that the difference between CFAA and non-CFAA sentencing seen in Table 21 is not due to discrepancies in offense level or criminal histories. The coefficient of -0.436 for the crime type of CFAA (as opposed to non-CFAA frauds) aligns with the descriptive observation that CFAA defendants receive fewer sentences that include prison time (with a fairly large effect size for a binary variable). The difference of about 2.7 months in the OLS model of sentence length also seems to be in line with the three-month difference between CFAA and non-CFAA sentences in Table 21. But this is a rudimentary model meant as a “sanity check” for the descriptive statistics; in addition to the simplifications mentioned above, there is a large difference between the number of CFAA crimes in the data (714, of which 266 had any prison sentence) and the number of non-CFAA fraud crimes (75,565 and 45,415).³⁹⁸ The results should therefore be read with caution.

Table 22: Regression results comparing CFAA and non-CFAA fraud sentences

	(1) <i>Prison?</i>	(2) <i>Months (if any)</i>
Crime Type (vs. Non-CFAA Fraud)		
CFAA	-0.436*** (0.053)	-2.740*** (0.709)
Offense level	0.137*** (0.001)	2.717*** (0.024)
Criminal history category	0.401*** (0.006)	5.057*** (0.059)
Demographic Controls	Y	Y
Constant	-1.280*** (0.029)	-28.026*** (0.462)
<i>N</i>	76,279	44,354
[Pseudo] R ²	0.283	0.572

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Model (1) shows the result of a probit regression on whether prison time was imposed. Model (2) shows the result of a linear regression on the length of sentences, in months, if a prison term was imposed. Demographic variables included age, sex, race, and education. Both regressions included only sentences with only one count of conviction.

³⁹⁷ The district with the most CFAA sentences in the USSC data set is the Central District of California, with 59 single-count CFAA sentences. There are 52 districts in the data set with 5 or fewer single-count CFAA convictions, including 15 with no CFAA convictions. Thus, for example, empirically comparing sentencing in the Northern District of California (46 single-count sentences) to Wyoming (1 single-count sentence) is not practical.

³⁹⁸ These numbers are lower than the total of 797 single-count CFAA crimes and 76,875 single-count non-CFAA fraud crimes because some records in the USSC data set were excluded because values for one or more of the control variables were missing.

4.3.6.3. Frequency of Sentencing Departures

Sentencing of different crime types can also be compared by looking at departures from the recommended ranges under the sentencing guidelines. Because the guidelines take offense levels, offender characteristics, and certain facts (e.g., amount of loss) into account when determining the presumptive sentencing range, comparing departures from those ranges controls (in a loose sense) for those variables. Departures may also show how often judges or prosecutors deem the guidelines not to reflect the true seriousness of a crime. An analysis of sentencing departures is, however, subject to biases from the guidelines' selection of factors that contribute to a crime's offense level and the offender's criminal history score.

Table 23 shows how often sentences departed from the guidelines for each type of crime (a "government sponsored" departure is a sentence that prosecutors proposed as below the guideline range, as opposed to a downward departure initiated by the judge). CFAA sentences were within the guidelines range 62.3% of the time, with virtually all departures being downward. Non-CFAA fraud crimes had a slightly higher rate of upward departures, but they also had more downward departures—and more of those were recommended by prosecutors. Burglary crimes tended to have more upward departures and fewer downward departures.

Table 23: Departures from the sentencing guideline ranges, by crime type

<i>Crime Type</i>	<i>Within Range</i>	<i>Upward</i>	<i>Gov't Sponsored</i>	<i>Downward</i>	<i>N</i>
CFAA	62.3%	0.8%	12.9%	24.1%	761
Non-CFAA Fraud	56.3%	1.9%	19.1%	22.7%	76,605
Burglary	66.4%	5.7%	10.6%	17.2%	679
Trespass	97.4%	2.6%	0.0%	0.0%	77

Note: The table shows the percentage of sentences that were within the range recommended by the guidelines or departed from those guidelines, either upward, downward at the government prosecutor's recommendation ("Gov't Sponsored"), or downward at the judge's discretion. This table includes only sentences for which there was a single count of conviction. Source: Combined USSC and PACER data set.

Whatever the reason for the difference between CFAA and non-CFAA fraud sentencing, it does not appear to be because judges are departing from recommended CFAA sentences more often. The opposite seems to be true: judges have followed the guideline recommendations more often with CFAA sentences than with non-CFAA fraud sentences. But departures from the guidelines for both crime types have overwhelmingly been to reduce sentences, not increase them.

4.4. Public Perceptions

To supplement the analysis of sentencing data, I conducted an empirical study of public attitudes about cybercrime. The goal of the study was to compare ratings of (a)(2) crimes to certain real-world crimes—specifically, trespass, burglary, and non-computer fraud crimes—using fact patterns drawn from crimes that are actually sentenced at the federal level.

4.4.1. Methodology

I conducted a between-subjects experiment in September 2018 that asked 499 participants to rate the seriousness, wrongfulness, and harmfulness of 28 short vignettes describing federal CFAA (a)(2), trespass, burglary, and fraud crimes. I used the seriousness, wrongfulness, and harmfulness metrics for consistency with previous work on perceptions of crime.³⁹⁹

Vignettes were based on the real-world scenarios described in Section 4.3. The vignettes are listed in Tables Table 24, Table 25, Table 26, and Table 27. Vignettes with multiple dollar amounts and vignette numbers were presented in two versions: one with a “low” loss of \$4,000 and another with a “high” loss of \$200,000.

Table 24: CFAA vignettes

#	Vignette	Off. Level
1	An IRS employee looks at a celebrity’s tax records out of curiosity.	8
2	A person installs monitoring software on another person’s computer without permission.	8
3	A person reads someone else’s e-mail without their permission.	8
4,5	An employee takes a copy of his employer’s confidential customer lists with him without permission when he quits. The customer list is worth \$4,000/\$200,000.	6/16
6,7	A business owner uses his customers’ passwords to log into a competitor’s web site to view the site design. The competitor spends \$4,000/\$200,000 investigating the incident.	6/16
8,9	A person downloads 100,000 email addresses from a company’s website with a security flaw. The company spends \$4,000/\$200,000 responding to the incident.	6/16

Table 25: Trespass vignettes

#	Vignette	Off. Level
10	A foreign national stows away on a cargo ship headed to the United States.	6
11	A passenger tries to bring a weapon through airport security.	8
12	A person camps in a national park where camping is not allowed.	4
13	A person knowingly trespasses in a wildlife refuge.	4
14	A person carrying a gun enters a fenced area of a nuclear power plant without permission.	8

Table 26: Burglary vignettes

#	Vignette	Off. Level
15	A person breaks into an unoccupied post office at night and steals \$200,000.	15
16	A person breaks into an unoccupied pharmacy and steals \$200,000 worth of drugs.	15
17	A person breaks into a unguarded railroad car and steals \$200,000 worth of goods.	15
18	A person breaks into an unoccupied bank at night and steals \$200,000.	15

³⁹⁹ See Mark Warr, *What is the Perceived Seriousness of Crimes?*, 27 CRIMINOLOGY 795, 796 (1989).

Table 27: Fraud vignettes

#	Vignette	Off. Level
19,20	A bank employee embezzles \$4,000/\$200,000 from his employer.	6/16
21,22	A postal employee steals mail containing \$4,000/\$200,000 in checks from dozens of homes	8/18
23,24	A scammer makes \$4,000/\$200,000 by sending letters that trick dozens of people into sending “processing fees” to receive fake lottery winnings	8/18
25,26	A person collects \$4,000/\$200,000 in Social Security payments by using someone else’s Social Security Number.	6/16
27,28	A pharmacist submits false Medicare claims totaling \$4,000/\$200,000.	6/16

The design uses each crime’s offense level under the U.S. Sentencing Guidelines as a control. As discussed in Section 3.2.1.2, the sentencing guidelines set a recommended sentence by cross-referencing the crime’s offense level, calculated based on the nature and circumstances of the crime, with an offender category that is determined by the offender’s criminal history. The vignettes were constructed to fall into one of two rough categories of offense levels. The “less serious” vignettes have offense levels of 4, 6, or 8.⁴⁰⁰ These levels correspond to a sentence of 0–6 months imprisonment for a first-time offender. The “more serious” vignettes were constructed to have an offense level of 15, 16, or 18, which allows CFAA, burglary, and fraud vignettes with the same dollar amounts to have roughly the same offense levels (the recommended sentence for an offense level of 16 would be 21–27 months for a first-time offender).

The survey instrument was organized as follows: After a consent page and instruction page, the next page asked participants to rate the seriousness of the 28 vignettes (presented in random order). The next page asked participants to rate the wrongfulness of the same 28 vignettes (presented in the same order as on the previous page). The following page asked for ratings of harmfulness. Demographic and summary questions followed.

The rating task was a slider for each vignette. The slider was labeled with “Not at all serious/harmful/wrongful” at one end and “Extremely serious/harmful/wrongful” on the other, but were not marked with intervals other than the endpoints and no numerical value was displayed. This was chosen in the hope that respondents would be able to make finer distinctions between the seriousness of crimes. Responses were recorded as integers between 0 (“not at all”) to 100 (“extremely”).

4.4.2. Results

I recruited 500 participants on Amazon Mechanical Turk. The survey was restricted to U.S. residents age 18 or over. After removing one response that said that the respondent was under 18, 499 responses remained. Survey participants were 54% male and 81% white. The median age was 35. Participants identified as 41% Democrat, 24% Republican, 29% Independent, and 2% other (3% chose not to answer; totals do not equal 100% due to rounding). Almost half of the respondents (47%) had at least a four-year college degree.

⁴⁰⁰ The offense level varies based on how the sentencing guidelines apply to the particular offense.

I ran extensive attention and quality checks on the responses. Because I asked each participant to rate versions of vignettes in which only the dollar value differed (e.g., embezzling \$4,000 versus \$200,000), I had a kind of built-in attention check; a careful rating of the vignettes should always rate the higher-loss version as at least as serious, wrongful, and harmful as the low-loss version. The high-loss version of the vignette thus loosely dominates the low-loss version. I call a vignette pair rated in the other direction an “inconsistently rated dominated vignette pair,” or IRDVP. IRDVPs could also be used as a quality check; in theory, the most reliable responses would have no IRDVPs.

As it turned out, only a small minority (68, 13.6%) had no IRDVPs at all. I believe that this is partly because of the granularity of the rating task measurement versus what participants perceive and partly because of the large number of dominated vignette pairs (24: 8 pairs rated three different times). When I loosened the criteria for an IRDVP to include only those where the difference in ratings is more than 10 points on the 100-point scale (i.e., to allow for ± 10 being roughly equivalent ratings), 44.4% of respondents had zero IRDVPs, 24.8% had one IRDVP, and 12.6% had two IRDVPs; 90.6% of all respondents had 3 or fewer IRDVPs.

Additionally, I analyzed IRDVPs based on the total difference between ratings (i.e., rating a \$4,000/\$200,000 vignette pair as 95/5 is a stronger signal of lack of response quality than 60/40 ratings). I also looked for ordering effects, distance effects (i.e., were vignettes more likely to be inconsistently rated if presented far apart?), and time taken completing the survey. I found some increase in IRDVPs as the pairs became more widely separated.

These results are robust at all the above levels of filtering based on IRDVPs. In fact, the more strict the attention-check criteria, the stronger the results appear to be—likely because the stricter checks filter out responses that tend to act as “noise.”

Figure 9 summarizes the responses (additional figures are provided in Appendix F). Crime ratings were generally in line with expectations, with burglary and fraud crimes receiving the highest ratings of seriousness, wrongfulness, and harmfulness. The higher-loss versions of vignettes received higher overall ratings than the lower-loss versions.

The ratings show that the vignettes included two very different types of federal trespass crimes. The scenarios involving camping in a national park or trespassing in a wildlife refuge were, as expected, judged to be among the least serious of the crimes described in the experiment. But “A passenger tries to bring a weapon through airport security” and “A person carrying a gun enters a fenced area of a nuclear power plant without permission” received fairly high ratings of seriousness, wrongfulness, and harmfulness. Stowing away on a cargo ship received a wide range of ratings but fell mostly in between the extremes of the other four vignettes. I speculate that the higher ratings for the airport and nuclear-plant trespass crimes may be due to the presence of weapons or because of the possible national security or terrorism implications of the vignettes.

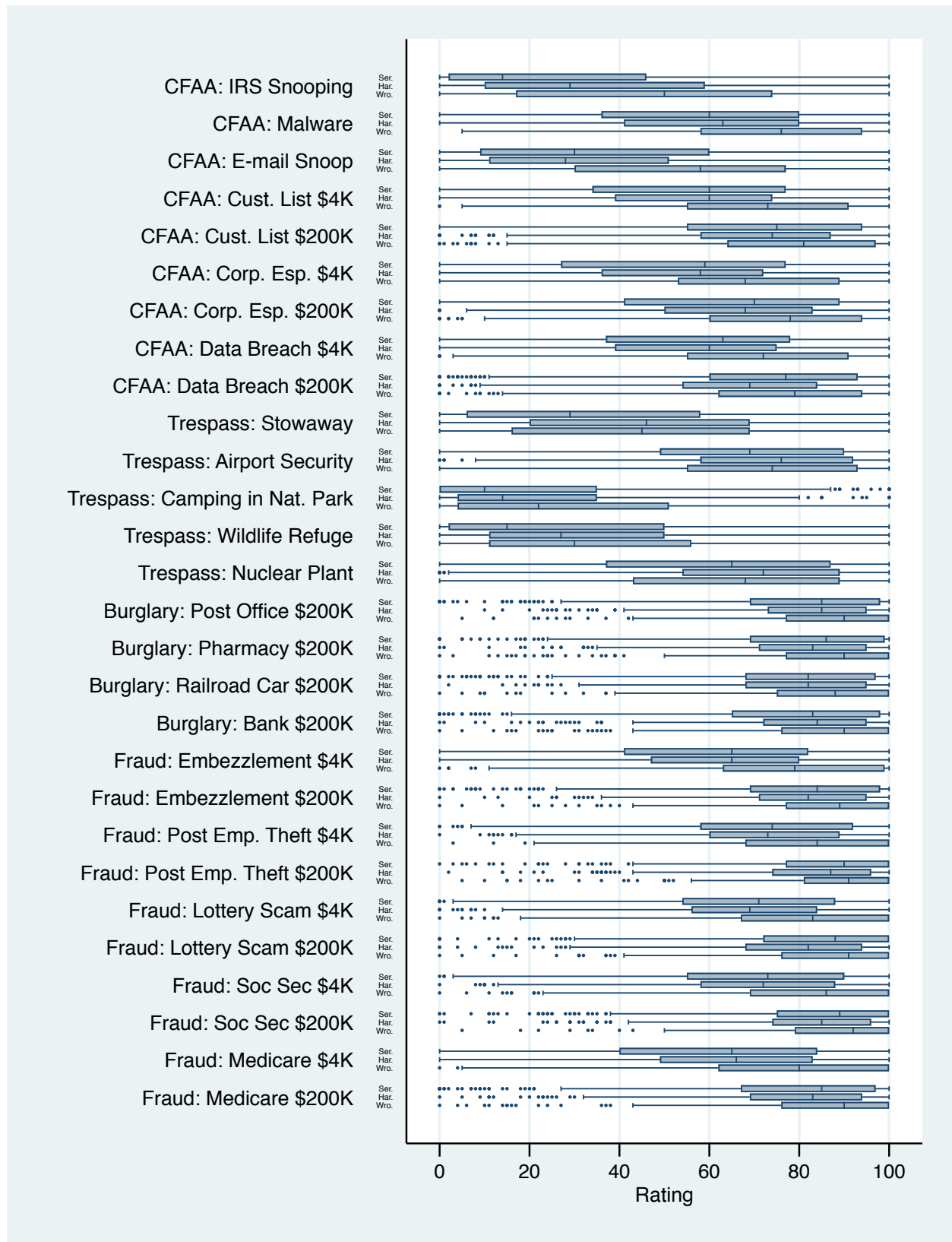


Figure 9: Summary of vignette responses

The ratings also show a wide range of responses on the CFAA crimes, suggesting a lack of consensus on the seriousness of cybercrime. Cybercrimes were generally rated to be less serious, wrongful, and harmful than burglary crimes, even at the same dollar value of loss. They were also rated as less serious, wrongful, and harmful than non-CFAA fraud crimes, even at the same dollar amounts of loss.

Table 28 shows the results of a random-effects model linear regression of the survey responses. The model includes interaction terms between offense level and crime type; many interaction categories are omitted because they do not appear in the sample (e.g., there are no trespass crimes with offense level 16 or 18) or because of collinearity (e.g., the only crimes with offense level 15 are burglary crimes, so offense level 15 is collinear with the crime type of burglary).

These results suggest that CFAA crimes are viewed as less serious than burglary or fraud crimes but as more serious than trespass crimes. They also reinforce the result that the vignettes included two types of trespass crimes: those that were relatively harmless (the ones with offense levels 4 and 6) and the serious ones involving airport security and nuclear plants (at offense level 8).

Table 28: Regression results for perceptions of crime types

	(1) Serious	(2) Wrongful	(3) Harmful
Crime Type (v. CFAA)			
Trespass	-15.907*** (0.915)	-25.208*** (1.150)	-16.033*** (1.075)
Burglary	37.287*** (0.961)	26.171*** (1.055)	37.177*** (1.288)
Fraud	12.823*** (0.665)	11.090*** (0.712)	11.464*** (0.844)
Offense Level (vs. 4)			
6	13.225*** (0.784)	10.594*** (0.942)	15.844*** (0.998)
8	13.217*** (0.965)	10.004*** (1.038)	17.293*** (1.146)
16	24.263*** (0.894)	16.223*** (1.020)	27.876*** (1.170)
18	24.606*** (0.985)	16.995*** (1.063)	31.246*** (1.210)
Crime Type x Off. Level			
Trespass x 6	4.822*** (1.409)	1.304 (1.481)	-4.369** (1.553)
Trespass x 8	29.876*** (1.404)	23.658*** (1.554)	20.749*** (1.606)
Fraud x 6	-3.468*** (0.574)	-2.417*** (0.553)	-3.735*** (0.620)
Demographics	Yes	Yes	Yes
Constant	47.039*** (4.641)	56.943*** (4.827)	41.392*** (5.645)
N	13,972	13,972	13,972
Clusters	499	499	499
R2 (Within)	0.454	0.407	0.399
R2 (Between)	0.052	0.056	0.052
R2 (Overall)	0.336	0.302	0.294

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Regression included demographic variables that are not listed in the table above; these results were not statistically significant.

4.5. Discussion

4.5.1. Policy Implications

These results support the idea that CFAA crimes are different from non-CFAA fraud crimes. But the results also suggest that trespass is not quite the right basis of comparison, either—at least not in terms of the federal trespass crime vignettes I used.

Court records show that there is a wide variety of crime sentenced under (a)(2), and that punishment varies across the different fact patterns. There are many cases involving the kind of “hacking” at which the CFAA was ostensibly aimed. But section (a)(2) has also been used extensively against employees in both government and the private sector who act outside the scope of their authority, and the sentences for this conduct have tended to be less harsh than those for the “hacking” cases—a fact which could support arguments either that the employee-misconduct cases do not belong in (a)(2) or that they are fine where they are, since it appears that these less-serious crimes are being punished less harshly. The question may be academic: sentences for employee misconduct declined in recent years and the Supreme Court’s recent *Van Buren* decision⁴⁰¹ is likely to further curtail—or even eliminate—future cases of this type.

These results add further support for the idea that CFAA crimes should have their own section of the sentencing guidelines. If (a)(2) computer crimes are unlike other fraud, then it makes little sense to calculate them under section 2B1.1 as if they were fraud. But neither would they fit the guidelines sections for burglary or trespass.

4.5.2. Limitations and Opportunities for Future Work

The results in this section are subject to several caveats and limitations. Perhaps foremost among these is the importance of prosecutorial discretion. Prosecutors have tremendous leeway in deciding whether to charge a crime, what crime(s) to charge, whether to plea bargain, what terms to offer in a plea bargain, and what sentences to request or recommend.⁴⁰² Thus, although court records and sentencing data will show the crimes that a defendant was convicted of, the data does not show what the defendant might have been charged with. The data is similarly limited by the fact that it was the result of a search for CFAA crimes: the same conduct charged as CFAA crimes in the data may have been charged as different crimes for different defendants. I have no way of knowing, based on this data, how often prosecutors chose to use other options instead of the CFAA.

Another limitation of this work is its analysis solely of federal crimes. This was necessary because state and federal crimes differ in so many ways—most importantly, for this work, in the different ways that sentences are determined at the federal and state levels—that comparing state and federal crimes to each other was impractical. But analyzing only federal

⁴⁰¹ See *Van Buren v. U.S.*, 593 U.S. ____ (2021).

⁴⁰² See, e.g., Robert L. Misner, *Recasting Prosecutorial Discretion*, 86 J. CRIM. L. & CRIMINOLOGY 717, 728–55 (1996) (discussing the prosecutor’s role in charging and plea bargaining).

crimes may skew the results toward large or noteworthy crimes or those that are inherently federal in nature (such as the IRS agents who looked at tax information without a valid purpose). It also limits the variety of crime fact patterns I can analyze; the trespass and burglary fact patterns are particularly limited in having to involve federal jurisdiction.⁴⁰³

Analyzing CFAA sentencing is also complicated by the fact that the CFAA often appears with other charges of conviction. When it does, it can be difficult to disentangle the effects of the various crimes on the final sentence. Where possible, I tried to account for this either by looking at cases where the CFAA was the only substantive charge (when making comparisons between CFAA cases) or by restricting analyses to cases with a single count of conviction (when comparing CFAA sentences to non-CFAA sentences).⁴⁰⁴ But this may also skew the results toward “simpler” conduct that would not be subject to multiple criminal statutes or multiple charges.

There is potential selection bias in the court records data. I believe, however, that the chance is minimal, given the high level of overlap between the court records data and the sentences in the Sentencing Commission data set. The merger of those two data sets may have introduced some bias, however, because certain fact patterns or judicial districts appeared to be overrepresented among sentences I was unable to cross-match.⁴⁰⁵

There may also be some subjectivity in the categorization of fact patterns in the sentencing analysis or in the selection of crimes for comparison in the attitudes study. The former was based on an overall impression of fact pattern commonalities after reviewing the court records, not on any previously determined taxonomy. Although I selected crime vignettes based on actual fact patterns, they still represent only a selection of the many different fact patterns for each crime; the survey experiment results are therefore limited to those fact patterns.

The analysis of fact patterns is subject to limits on the ability to know the “ground truth” of any case. Court records will state the facts as alleged, as pled to, and sometimes as proven at trial, but rarely do these provide the information needed to know all the facts of a case. Furthermore, some of the court records I use as data sources are works of advocacy and are thus prone to presenting the facts from a certain viewpoint. I attempted to use the most authoritative sources available when reviewing case facts.⁴⁰⁶

Information about sentences is limited by choices in coding the data. For example, I entered into the data set only the final charges of conviction, not statutes that were charged but later dropped or dismissed, because only those charges subject to a guilty verdict or guilty plea can be considered to be proven. But the court records I downloaded do contain information about the statutes that were originally charged; they could be used in future research comparing initial charging decisions to final charges. For example, future research could look at charged

⁴⁰³ See *supra* Parts 4.3.3 and 4.3.4.

⁴⁰⁴ See *supra* notes 289 and 390.

⁴⁰⁵ See *supra* note 256.

⁴⁰⁶ See *supra* Section 4.3.1.

and pled statutes to evaluate whether the (a)(2) is frequently used to “plead out” from crimes subject to harsher punishments.

The use of public perceptions in evaluating whether criminal sentences are “fair” is subject to the same caveats outlined in Chapter 3.⁴⁰⁷ Although public opinion can certainly inform sentencing policy, it would be a mistake to impose criminal sentences purely according to public opinion.

As mentioned in the previous section, the results say how computer crime is *not* perceived or punished. But the question of what cybercrime *is* remains open. Future work could compare computer crimes to other types of crimes. In particular, it would be interesting to see how perceptions of (a)(2) access-to-information crimes compare to wiretap crimes,⁴⁰⁸ which are sentenced as “Privacy and Eavesdropping” crimes under section 2H3.1 of the guidelines, or Stored Communications Act crimes,⁴⁰⁹ which, like CFAA crimes, are sentenced as fraud under section 2B1.1.

Another area for future work could be to retrieve court documents and analyze fact patterns for all CFAA cases, not just (a)(2) cases. I expect that most of the variation among fact patterns falls under (a)(2), but it could be interesting to see, for example, whether (a)(4) and (a)(5) are primarily unauthorized-access cases or also have a large number of cases in which the offender exceeded authorized access. The manual method I used to review sentencing documents presents a potential obstacle to extending this chapter’s analysis to other crime types, so it could also be interesting to see whether those documents are amenable to automated analysis.

4.5.3. Conclusion

To the question posed at the start of this chapter—“What *is* computer crime, really?”—the results provide a partial answer. Although the elements of (a)(2) crimes are sentenced under the guidelines provisions for fraud crimes and often prompt analogies to trespass or burglary crimes, the analogies do not extend to actual sentencing patterns or to public perceptions of the seriousness of the crimes. CFAA computer crimes are sentenced differently than trespass, burglary, or non-CFAA fraud crimes, and sentencing even varies between different types of CFAA (a)(2) crimes. And public perceptions of cybercrime appear to be different from those for trespass, burglary, and other frauds as well, at least for the vignettes used in this study.

The essential nature of computer crime in an absolute or philosophical sense remains open to debate. But if the public views computer crimes as something fundamentally different from other crimes, arguing by analogy may not be the best approach for analyzing how computer crimes should be punished. And if CFAA crimes are sentenced differently from other crimes, as shown here, that could lend support to the argument that these crimes are a poor fit for the fraud

⁴⁰⁷ See *supra* Section 3.5.3.

⁴⁰⁸ See 18 U.S.C. §§ 2511–2523.

⁴⁰⁹ See 18 U.S.C. §§ 2701–2713.

section of the guidelines—or it could undermine that argument by showing that the CFAA’s treatment as a fraud crime for sentencing purposes is not enough to prevent it from being sentenced differently from other fraud crimes in practice.

5. Conclusion

The previous chapters explored some of the consequences or perceptions of cybercrime: the economics of credit card reissue after a breach, and whether cybercrime punishments are aligned with public attitudes about those crimes. The research also raises issues relating to the uncertainty in public data about data breach and credit card misuse, the fundamental nature of cybercrime, and how access-to-information computer crimes are prosecuted and sentenced.

Chapter 2 shows that reissuing cards immediately after a breach appears to be less costly than waiting for attempted fraud before reissuing. Despite uncertainty in the data, this result is fairly robust, with an estimated 9% probability that waiting to reissue cards until fraud is detected would save money. But uncertainty in the data highlights a larger policy problem. Despite extensive breach reporting requirements, information about breaches is often incomplete. Better information about the causes, effects, scope, and frequency of data breach could improve data security decision making. Resources could usefully be targeted to getting better data for variables such as (in the context of credit card breaches) how many card records are breached each year, how identity thieves get access to card data, and how effective fraud monitoring is at preventing card misuse. A handful of states require reporting detailed information to state attorneys general; a similar requirement at the federal level could improve analyses of breaches. More fundamentally, the uncertainty in information about data breaches highlights the ways in which data may be flawed. Data might not exist, it could be of poor quality, or it could exist but be used incorrectly. Policy decisions that rely on uncertain data should take these different types of uncertainty into account.⁴¹⁰

The analysis of cybercrime attitudes in Chapter 3 suggests that CFAA sentences are misaligned with the public perceptions. The amount of loss has an outsized effect on recommended sentences compared to the importance of that factor on perceptions of crime seriousness. But an attacker's motive and the sensitivity of the data, which have a large effect on perceptions of seriousness, have only minimal effects on calculations under the guidelines. These results provide empirical support for arguments that CFAA sentencing is miscategorized in the federal sentencing guidelines.

Chapter 4's examination of sentencing data and public perceptions of the nature of cybercrimes has a few policy implications. First, it reinforces the idea that CFAA (a)(2) crimes should not be sentenced according to the fraud section of the sentencing guidelines. Second, the breadth of fact patterns sentenced under (a)(2) suggests that it may have been overused as an easy statute to punish any bad behavior that uses a computer. The data shows that this use (which some might call misuse) has been in decline over the years—a trend that should accelerate in the wake of *Van Buren*.⁴¹¹ Third, the wide variation in ratings of computer crimes

⁴¹⁰ See James T. Graves, Alessandro Acquisti & Nicolas Christin, *Big Data and Bad Data: On the Sensitivity of Security Policy to Imperfect Information*, 83 U. CHICAGO L. REV. 117 (2016).

⁴¹¹ See *Van Buren v. U.S.*, 593 U.S. ____ (2021).

in the survey experiment suggests that there is a lack of consensus about the seriousness of cybercrimes; this lack of consensus could make reforming the CFAA more difficult.

Appendix A. U.S. Sentencing Guidelines Sentencing Table

Table 29: U.S. Sentencing Guidelines Sentencing Table

	Offense Level	Criminal History Category (Criminal History Points)					
		I (0 or 1)	II (2 or 3)	III (4, 5, 6)	IV (7, 8, 9)	V (10, 11, 12)	VI (13 or more)
Zone A	1	0-6	0-6	0-6	0-6	0-6	0-6
	2	0-6	0-6	0-6	0-6	0-6	1-7
	3	0-6	0-6	0-6	0-6	2-8	3-9
	4	0-6	0-6	0-6	2-8	4-10	6-12
	5	0-6	0-6	1-7	4-10	6-12	9-15
	6	0-6	1-7	2-8	6-12	9-15	12-18
	7	0-6	2-8	4-10	8-14	12-18	15-21
	8	0-6	4-10	6-12	10-16	15-21	18-24
Zone B	9	4-10	6-12	8-14	12-18	18-24	21-27
	10	6-12	8-14	10-16	15-21	21-27	24-30
	11	8-14	10-16	12-18	18-24	24-30	27-33
Zone C	12	10-16	12-18	15-21	21-27	27-33	30-37
	13	12-18	15-21	18-24	24-30	30-37	33-41
	14	15-21	18-24	21-27	27-33	33-41	37-46
Zone D	15	18-24	21-27	24-30	30-37	37-46	41-51
	16	21-27	24-30	27-33	33-41	41-51	46-57
	17	24-30	27-33	30-37	37-46	46-57	51-63
	18	27-33	30-37	33-41	41-51	51-63	57-71
	19	30-37	33-41	37-46	46-57	57-71	63-78
	20	33-41	37-46	41-51	51-63	63-78	70-87
	21	37-46	41-51	46-57	57-71	70-87	77-96
	22	41-51	46-57	51-63	63-78	77-96	84-105
	23	46-57	51-63	57-71	70-87	84-105	92-115
	24	51-63	57-71	63-78	77-96	92-115	100-125
	25	57-71	63-78	70-87	84-105	100-125	110-137
	26	63-78	70-87	78-97	92-115	110-137	120-150
	27	70-87	78-97	87-108	100-125	120-150	130-162
	28	78-97	87-108	97-121	110-137	130-162	140-175
	29	87-108	97-121	108-135	121-151	140-175	151-188
	30	97-121	108-135	121-151	135-168	151-188	168-210
	31	108-135	121-151	135-168	151-188	168-210	188-235
	32	121-151	135-168	151-188	168-210	188-235	210-262
	33	135-168	151-188	168-210	188-235	210-262	235-293
	34	151-188	168-210	188-235	210-262	235-293	262-327
	35	168-210	188-235	210-262	235-293	262-327	292-365
	36	188-235	210-262	235-293	262-327	292-365	324-405
	37	210-262	235-293	262-327	292-365	324-405	360-life
	38	235-293	262-327	292-365	324-405	360-life	360-life
	39	262-327	292-365	324-405	360-life	360-life	360-life
	40	292-365	324-405	360-life	360-life	360-life	360-life
	41	324-405	360-life	360-life	360-life	360-life	360-life
	42	360-life	360-life	360-life	360-life	360-life	360-life
	43	life	life	life	life	life	life

Appendix B. Regression Tables for the Chapter 3 Between-Subjects Experiments

Table 30: Ordered probit marginal effects for the Type of Data experiment

	Wrongful	Harmful	Serious	Harsh	Sensitive	Respons.	Clever
Medical data	-0.104 (0.142)	0.194 (0.145)	0.076 (0.148)	-0.028 (0.145)	0.970*** (0.151)	0.015 (0.153)	0.008 (0.143)
Female	0.435** (0.147)	0.259 (0.156)	0.348* (0.153)	0.084 (0.158)	0.349* (0.162)	0.416** (0.150)	-0.027 (0.143)
US birth	-0.209 (0.227)	0.141 (0.348)	0.322 (0.295)	0.040 (0.293)	0.521 (0.326)	-0.354 (0.447)	0.177 (0.214)
CFIP score	0.563*** (0.110)	0.197 (0.116)	0.304** (0.104)	0.295** (0.102)	0.501*** (0.117)	0.281* (0.116)	0.235 (0.131)
Freq. aff by cybercrime	-0.016 (0.132)	-0.003 (0.110)	-0.127 (0.117)	-0.081 (0.130)	-0.142 (0.131)	0.122 (0.095)	-0.301* (0.131)
Fake personal info	-0.010 (0.061)	-0.032 (0.060)	-0.046 (0.056)	-0.093 (0.056)	-0.083 (0.063)	0.040 (0.061)	-0.018 (0.058)
Media awareness	-0.083 (0.051)	-0.026 (0.049)	0.009 (0.049)	0.016 (0.044)	0.033 (0.051)	0.077 (0.051)	0.064 (0.053)
AC: Data	0.490 (0.257)	0.396 (0.297)	0.117 (0.252)	-0.106 (0.248)	0.069 (0.282)	-0.660* (0.307)	-0.204 (0.239)
AC: Context	-0.287 (0.166)	-0.296 (0.179)	-0.474** (0.164)	-0.306* (0.150)	-0.343* (0.163)	0.100 (0.173)	0.154 (0.165)
AC: Scope	-0.379 (0.194)	-0.564** (0.209)	-0.502* (0.212)	-0.165 (0.184)	-0.410 (0.229)	0.256 (0.196)	-0.127 (0.208)
<i>N</i>	239	239	239	239	239	239	239
pseudo <i>R</i> ²	0.079	0.048	0.053	0.047	0.128	0.060	0.045

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The table shows ordered probit regression results for responses to the seven main Likert questions in the Type of Data experiment. The “Medical data” condition is versus the baseline condition of directory data.

Regressions also included categorical control variables for occupation, age, education, and work situation.

Table 31: Ordered probit regression results for the Scope experiment

	Wrongful	Harmful	Serious	Harsh	Sensitive	Respons.	Clever
log(Num. Records)	0.070** (0.027)	0.078** (0.026)	0.159*** (0.028)	0.107*** (0.026)	0.135*** (0.031)	0.064* (0.026)	0.057* (0.025)
Female	0.186 (0.097)	0.045 (0.095)	-0.014 (0.095)	0.109 (0.092)	0.240* (0.110)	-0.145 (0.094)	0.096 (0.093)
US birth	-0.249 (0.194)	0.028 (0.211)	-0.296 (0.159)	-0.309 (0.207)	-0.210 (0.272)	-0.033 (0.234)	-0.435 (0.234)
CFIP score	0.361*** (0.067)	0.242*** (0.070)	0.381*** (0.071)	0.241*** (0.067)	0.628*** (0.081)	0.210** (0.069)	0.261*** (0.065)
Freq. aff by cybercrime	-0.095 (0.063)	-0.072 (0.060)	-0.187** (0.063)	-0.102 (0.062)	-0.185* (0.076)	-0.023 (0.061)	-0.025 (0.063)
Fake personal info	0.049 (0.040)	-0.045 (0.038)	-0.019 (0.038)	-0.013 (0.037)	-0.032 (0.045)	0.017 (0.039)	0.063 (0.039)
Media awareness	-0.044 (0.032)	-0.028 (0.032)	-0.032 (0.031)	-0.027 (0.029)	-0.036 (0.036)	0.047 (0.033)	-0.006 (0.031)
AC: Data	-0.020 (0.138)	-0.133 (0.137)	-0.173 (0.134)	-0.121 (0.134)	0.038 (0.162)	0.010 (0.139)	0.334* (0.142)
AC: Context	0.028 (0.136)	-0.010 (0.144)	-0.082 (0.119)	-0.159 (0.132)	0.276 (0.162)	0.107 (0.131)	0.163 (0.141)
AC: Scope	0.104 (0.126)	-0.031 (0.133)	0.072 (0.127)	0.262* (0.130)	0.030 (0.151)	-0.056 (0.128)	0.216 (0.140)
<i>N</i>	583	583	583	583	583	583	583
pseudo R^2	0.048	0.029	0.046	0.034	0.097	0.023	0.031

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The table shows ordered probit regression results for responses to the seven main Likert questions in the Scope experiment. Regressions also included categorical control variables for occupation, age, education, and work situation.

Table 32: Ordered probit regression results for the Motivation experiment (vs. Profiteer)

	Wrongful	Harmful	Serious	Harsh	Pot. Harm	Sensitive	Respons.	Clever
Student	-0.878*** (0.151)	-0.327* (0.148)	-0.596*** (0.150)	-0.793*** (0.145)	-0.051 (0.150)	0.201 (0.141)	0.034 (0.141)	0.217 (0.147)
Activist	-0.795*** (0.150)	-0.279 (0.145)	-0.538*** (0.152)	-0.497*** (0.147)	-0.294 (0.159)	0.130 (0.154)	0.100 (0.145)	0.191 (0.152)
Female	0.035 (0.121)	-0.037 (0.123)	0.056 (0.126)	-0.051 (0.128)	0.068 (0.129)	-0.106 (0.121)	0.364** (0.124)	0.001 (0.119)
US birth	-0.088 (0.212)	0.078 (0.259)	-0.050 (0.225)	0.335 (0.252)	0.042 (0.274)	-0.268 (0.234)	0.053 (0.318)	-0.339 (0.247)
CFIP score	0.238** (0.090)	0.181 (0.094)	0.295** (0.092)	0.223* (0.097)	0.255** (0.088)	0.341*** (0.087)	0.140 (0.087)	0.371*** (0.085)
Freq. aff by cybercrime	0.084 (0.093)	-0.047 (0.085)	0.114 (0.092)	-0.014 (0.098)	0.050 (0.091)	0.011 (0.095)	-0.121 (0.090)	-0.044 (0.095)
Fake personal info	0.003 (0.053)	-0.007 (0.051)	0.052 (0.053)	-0.007 (0.052)	0.027 (0.052)	-0.029 (0.052)	0.059 (0.051)	-0.045 (0.053)
Media awareness	0.009 (0.044)	0.100* (0.047)	0.053 (0.045)	0.033 (0.043)	0.100* (0.047)	0.026 (0.042)	0.030 (0.044)	-0.029 (0.042)
AC: Data	-0.313** (0.121)	-0.115 (0.131)	-0.220 (0.121)	-0.285* (0.128)	-0.223 (0.138)	-0.510*** (0.126)	0.081 (0.130)	-0.002 (0.135)
AC: Context	0.058 (0.155)	0.205 (0.151)	0.031 (0.159)	0.093 (0.155)	-0.170 (0.157)	0.032 (0.156)	0.250 (0.160)	0.192 (0.159)
AC: Scope	0.039 (0.138)	-0.113 (0.129)	0.091 (0.133)	-0.042 (0.139)	-0.079 (0.142)	0.110 (0.135)	-0.079 (0.130)	0.205 (0.142)
AC: Motivation	-0.208 (0.179)	-0.140 (0.192)	-0.244 (0.177)	-0.234 (0.170)	-0.327 (0.188)	-0.567** (0.178)	-0.126 (0.174)	-0.014 (0.189)
<i>N</i>	361	361	361	361	361	361	361	361
pseudo R^2	0.083	0.046	0.052	0.071	0.057	0.056	0.033	0.048

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The table shows ordered probit regression results for responses to the eight main Likert questions in the Motivation experiment. The “Student” and “Activist” motivation conditions are versus the “Profiteer” baseline condition. Regressions also included categorical control variables for occupation, age, education, and work situation.

Table 33: Ordered probit regression results for the Consequences experiment (vs. Low)

	Wrongful	Harmful	Serious	Harsh	Pot. Harm	Sensitive	Respons.	Clever
Acme High	0.179 (0.123)	0.407*** (0.122)	0.083 (0.119)	0.338** (0.123)	0.147 (0.137)	-0.009 (0.140)	-0.123 (0.116)	-0.020 (0.118)
Customers High	0.042 (0.125)	0.377** (0.120)	0.131 (0.121)	0.236* (0.118)	0.093 (0.138)	0.040 (0.151)	0.112 (0.126)	-0.125 (0.124)
Female	0.157 (0.106)	0.113 (0.103)	0.163 (0.101)	0.150 (0.101)	0.261* (0.116)	0.201 (0.122)	0.129 (0.106)	0.089 (0.103)
US birth	0.067 (0.241)	-0.116 (0.216)	0.157 (0.241)	0.116 (0.240)	0.008 (0.218)	-0.130 (0.269)	0.071 (0.287)	-0.096 (0.213)
CFIP score	0.212** (0.076)	0.168* (0.082)	0.294*** (0.078)	0.167* (0.080)	0.417*** (0.101)	0.650*** (0.104)	0.222** (0.074)	0.119 (0.078)
Freq. aff by cybercrime	-0.021 (0.079)	-0.002 (0.076)	-0.034 (0.078)	0.007 (0.077)	-0.015 (0.088)	-0.099 (0.097)	0.010 (0.075)	0.020 (0.073)
Fake personal info	-0.108* (0.043)	-0.054 (0.043)	-0.094* (0.043)	-0.115** (0.041)	-0.026 (0.052)	0.000 (0.048)	0.091* (0.043)	0.017 (0.041)
Media awareness	-0.047 (0.039)	0.028 (0.037)	-0.029 (0.039)	-0.023 (0.039)	0.075 (0.045)	0.052 (0.048)	0.042 (0.040)	0.028 (0.037)
AC: Data	0.416** (0.155)	0.139 (0.140)	0.243 (0.144)	0.211 (0.143)	0.327* (0.167)	0.471* (0.186)	0.326* (0.143)	0.090 (0.148)
AC: Context	-0.090 (0.128)	-0.068 (0.114)	0.039 (0.126)	-0.060 (0.119)	-0.054 (0.140)	-0.013 (0.149)	-0.336** (0.125)	-0.128 (0.129)
AC: Scope	-0.010 (0.109)	-0.111 (0.111)	-0.104 (0.111)	-0.127 (0.109)	-0.070 (0.124)	0.119 (0.133)	0.110 (0.115)	0.145 (0.108)
AC: Consequence	-0.089 (0.166)	-0.121 (0.202)	-0.130 (0.185)	-0.206 (0.186)	-0.183 (0.213)	-0.175 (0.205)	0.197 (0.165)	0.068 (0.179)
<i>N</i>	479	479	479	479	479	479	479	479
pseudo R^2	0.047	0.034	0.040	0.040	0.078	0.117	0.034	0.017

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The table shows ordered probit regression results for responses to the eight main Likert questions in the Consequences experiment. The “Acme High” and “Customers High” motivation conditions are the conditions in which Acme was described as experiencing high losses and its customers were described as experiencing high losses, respectively. Both were rare versus the “Low” baseline condition in which Acme was described as experiencing minimal losses. Regressions also included categorical control variables for occupation, age, education, and work situation.

Table 34: Ordered probit regressions for the Co-Responsibility experiment

	Wrongful	Harmful	Serious	Harsh	Pot. Harm	Sensitive	Respons.	Clever
Not Patched	0.133 (0.136)	0.102 (0.136)	0.157 (0.133)	0.074 (0.132)	0.087 (0.151)	-0.370* (0.164)	0.423*** (0.128)	-0.184 (0.136)
Female	0.197 (0.147)	0.192 (0.153)	0.144 (0.151)	0.060 (0.143)	0.154 (0.162)	-0.045 (0.175)	0.151 (0.142)	0.029 (0.149)
US birth	0.225 (0.356)	-0.758* (0.298)	0.366 (0.274)	0.227 (0.234)	-0.337 (0.433)	-0.593 (0.456)	-0.509 (0.356)	0.214 (0.391)
CFIP score	0.576*** (0.120)	0.391** (0.130)	0.557*** (0.117)	0.385*** (0.113)	0.701*** (0.137)	1.087*** (0.141)	0.249* (0.113)	0.364** (0.125)
Freq. aff by cybercrime	-0.007 (0.084)	0.035 (0.089)	-0.026 (0.104)	-0.048 (0.094)	0.034 (0.107)	0.011 (0.118)	0.097 (0.100)	0.002 (0.098)
Fake personal info	-0.016 (0.059)	-0.154* (0.065)	0.001 (0.066)	-0.121 (0.063)	0.025 (0.070)	-0.004 (0.066)	-0.051 (0.062)	0.066 (0.070)
Media awareness	0.030 (0.048)	0.113* (0.050)	0.093 (0.048)	0.076 (0.048)	0.041 (0.056)	-0.069 (0.062)	0.177** (0.054)	0.064 (0.054)
AC: Data	-0.271 (0.206)	-0.305 (0.219)	-0.343 (0.189)	-0.060 (0.218)	-0.071 (0.206)	-0.202 (0.270)	-0.197 (0.202)	0.185 (0.201)
AC: Context	-0.359* (0.162)	-0.359* (0.150)	-0.286 (0.161)	-0.234 (0.148)	-0.553** (0.178)	-0.251 (0.192)	-0.144 (0.162)	0.075 (0.169)
AC: Scope	0.007 (0.189)	0.234 (0.177)	0.271 (0.171)	0.082 (0.160)	0.494** (0.192)	0.384* (0.177)	0.226 (0.181)	0.200 (0.169)
AC: Patched	-0.294 (0.234)	-0.333 (0.229)	-0.184 (0.209)	-0.283 (0.212)	-0.277 (0.286)	-0.462 (0.284)	-0.359 (0.225)	0.032 (0.240)
<i>N</i>	276	276	276	276	276	276	276	276
pseudo R^2	0.061	0.053	0.052	0.039	0.107	0.167	0.057	0.050

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The table shows ordered probit regression results for responses to the eight main Likert questions in the Co-responsibility experiment. The “Not Patched” condition is versus the “Patched” baseline condition in which Acme was described as having patched its servers. Regressions also included categorical control variables for occupation, age, education, and work situation.

Table 35: Ordered probit regressions for the Context experiment (vs. Bank)

	Wrongful	Harmful	Serious	Harsh	Pot. Harm	Sensitive	Respons.	Clever
Government	-0.055 (0.119)	0.013 (0.121)	-0.027 (0.125)	-0.030 (0.116)	0.147 (0.139)	-0.121 (0.142)	0.152 (0.118)	-0.023 (0.116)
Non-Profit	0.048 (0.123)	-0.029 (0.124)	-0.222 (0.122)	0.030 (0.121)	0.099 (0.140)	-0.208 (0.155)	-0.361** (0.120)	-0.185 (0.121)
Org. size	0.055 (0.044)	0.064 (0.041)	0.045 (0.043)	0.053 (0.043)	0.133** (0.048)	0.148** (0.050)	0.059 (0.046)	0.142** (0.046)
Female	0.002 (0.102)	0.000 (0.101)	-0.044 (0.100)	-0.018 (0.099)	0.090 (0.116)	0.068 (0.118)	0.157 (0.096)	0.127 (0.100)
US birth	-0.069 (0.281)	-0.094 (0.276)	0.071 (0.226)	-0.157 (0.250)	-0.292 (0.284)	0.158 (0.376)	0.116 (0.301)	-0.050 (0.276)
CFIP score	0.354*** (0.073)	0.191* (0.077)	0.376*** (0.073)	0.207** (0.080)	0.405*** (0.085)	0.518*** (0.077)	0.139 (0.075)	0.135 (0.073)
Freq. aff by cybercrime	-0.020 (0.064)	-0.027 (0.063)	-0.052 (0.064)	-0.044 (0.060)	-0.118 (0.073)	0.004 (0.077)	-0.026 (0.064)	0.047 (0.069)
Fake personal info	-0.021 (0.041)	0.003 (0.040)	-0.016 (0.040)	-0.007 (0.037)	0.053 (0.045)	-0.078 (0.044)	0.013 (0.042)	-0.009 (0.040)
Media awareness	-0.026 (0.036)	-0.046 (0.038)	0.030 (0.036)	-0.010 (0.035)	-0.030 (0.044)	-0.010 (0.043)	0.065 (0.037)	0.062 (0.037)
AC: Data	0.023 (0.153)	0.019 (0.151)	0.029 (0.153)	0.017 (0.152)	0.376* (0.161)	0.372* (0.156)	-0.184 (0.123)	-0.053 (0.142)
AC: Context	-0.003 (0.129)	0.066 (0.133)	-0.196 (0.128)	0.010 (0.134)	-0.024 (0.144)	-0.036 (0.160)	-0.101 (0.123)	-0.153 (0.129)
AC: Scope	-0.152 (0.122)	0.051 (0.136)	-0.101 (0.126)	-0.035 (0.123)	-0.028 (0.144)	0.168 (0.142)	-0.022 (0.127)	0.035 (0.126)
N	502	502	502	502	502	502	502	502
pseudo R^2	0.044	0.022	0.045	0.029	0.073	0.092	0.028	0.034

Notes: The table shows ordered probit regression results for responses to the eight main Likert questions in the Context experiment. The “Government” and “Non-profit” conditions are versus the “Bank” baseline condition. Regressions also included categorical control variables for occupation, age, education, and work situation.

Appendix C. Inter-Respondent Heterogeneity in the Chapter 3 Factorial Experiments

To explore the extent to which respondents agree in their perceptions, I ran individual regressions for each of the 223 respondents in the experiment. The statistical power in the individual-level regressions is limited by the fact that each respondent rated only 25 vignettes.⁴¹² The explanatory power of many of the individual regressions is reasonably good, however. Adjusted R^2 values range from -0.305 to .952 with a median of 0.627.

Table 36 lists the percentage of responses for which each coefficient was statistically significant at $p < 0.05$ and $p < 0.01$. As should be expected from such a small value of N , only a small percentage of individual regressions showed statistically significant coefficients. The most frequently significant coefficient (other than the constant term) is $\log(\text{Sentence})$, which was significant at in 34% of individual regressions. All the variables of interest except those involving organization type were significant at rates higher than the corresponding level (i.e., the coefficient was significant at $p < 0.05$ for more than 5% of responses).

Table 37 shows summary statistics for the coefficients across individual-level regressions. Figure 10 and Figure 11 are histograms of the coefficients for each variable of interest across the individual-level regressions. As the table and figures show, there is wide variation in the coefficients that result from individual-level regressions. Unsurprisingly, the distributions are often skewed in the same direction as overall-level results, but each factor seems to have both negative and positive correlations with perceived severity depending on the respondent. But note that this table summarizes coefficients for all regression results regardless of whether the coefficients it summarizes are statistically significant.

These results suggest—though not conclusively, considering the small number of observations per respondent—that there is quite a bit of variation in how individuals weigh different factors of cybercrime.

⁴¹² As discussed in Section 3.4.1.2, *supra*, I scaled back to 25 vignettes per respondent after a pilot study with 40 vignettes per person exhibited technical problems and high dropout rates.

Table 36: Statistically-significant coefficients as percentages of individual-level regressions

	% $p < 0.01$	% $p < 0.05$
log(Records)	4.0	12.1
log(Org Loss)	2.7	10.3
log(Cust Loss)	1.3	8.1
Organization (vs. Bank)		
Government	0.9	5.8
Non-profit	0.9	7.2
Insurer	1.3	4.5
Data (vs. E-mail)		
Name, addr, SSN	5.8	14.8
Health	6.3	13.0
Name, phone, addr, DOB, SSN	7.2	16.1
Name, phone, addr	1.3	9.4
Name, user ID, pwd	3.1	10.8
Motivation (vs. Profiteer)		
Student	6.7	15.2
Activist	9.0	16.1
log(Sentence)	13.0	33.6
Probation	5.4	12.6
log(Sentence) x Probation	2.7	12.6
_cons	22.0	41.7

Notes: The table shows the percentage of individual-level regressions with statistically significant coefficients for each variable. For example, the coefficient for log (Records) was statistically significant at $p < 0.05$ for 12.1% of the individual-level regressions.

Table 37: Summary statistics for coefficients across individual-level regressions

var	mean	sd	5%	median	95%	N
cons	56.29	42.44	-14.77	54.57	130.17	223
log(records)	0.66	1.43	-1.51	0.51	3.23	223
log(cust_loss)	0.72	4.22	-5.66	0.31	8.09	223
log(org_loss)	0.52	1.86	-2.57	0.40	3.43	223
Organization (vs. Bank)						
Govt.	-1.19	16.57	-31.70	-2.03	24.74	223
Non-Profit	-2.76	16.35	-27.97	-2.01	21.63	223
Insurer	-1.10	16.04	-25.12	-0.19	23.11	223
Data (vs. E-mail)						
Name, addr, SSN	9.25	21.39	-19.21	6.82	44.14	223
Health	10.63	24.10	-31.56	9.77	49.37	217
Directory + DOB, SSN	11.93	21.61	-18.72	9.55	48.90	223
Directory	5.53	20.21	-27.15	4.03	40.15	222
Name, user ID, password	7.69	19.35	-24.14	5.45	43.34	221
Motive (vs. Profiteer)						
Student	-8.94	16.44	-34.79	-7.71	13.03	223
Activist	-10.03	16.18	-37.59	-8.91	13.51	223
log(sentence)	-9.01	9.61	-23.21	-9.29	8.36	223
probation	6.68	32.42	-49.98	6.12	57.09	223
log(sentence) x probation	2.58	11.94	-19.37	3.26	21.98	223

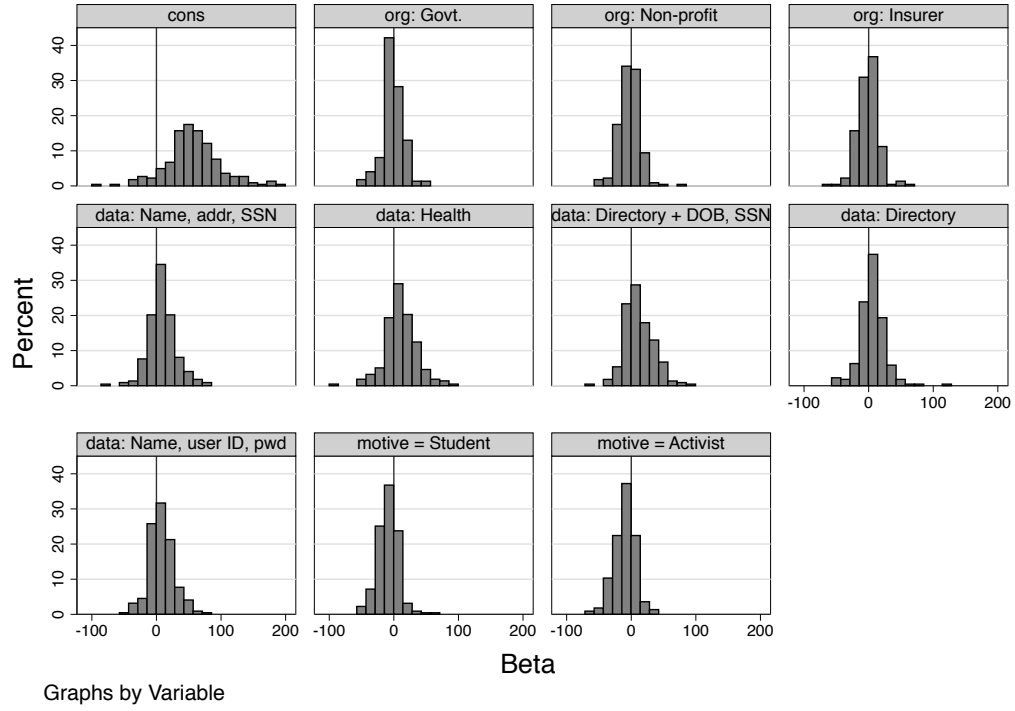


Figure 10: Distribution of β values over respondent-level models for non-log-scaled variables of interest

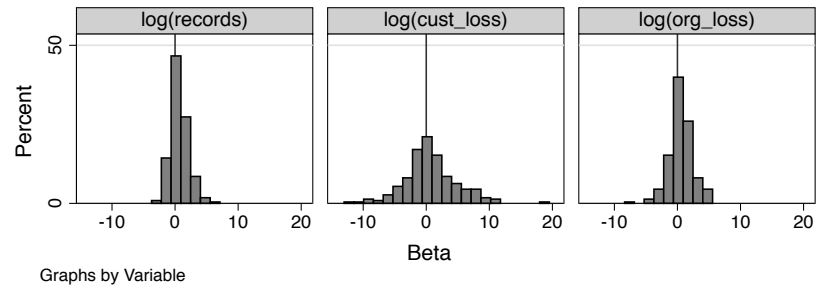


Figure 11: Distribution of β values over respondent-level models for log-scaled variables of interest

Appendix D. Example Survey Text for the Chapter 3 Between-Subjects Experiments

The six between-subjects cybercrime attitudes experiments were similar but not identical. All experiments used the same introduction, privacy attitude and demographic questions, and general structure. Each experiment had specific scenario text and attention-check questions that asked about the manipulation in that experiment's scenario. The closing open-ended questions also varied slightly between experiments. The question about the *potential* consequences of Tom Smith's actions was added for the Motivation, Consequences, Co-Responsibility, and Context experiments.

To save space, the survey text from one of the six experiments (the Motivation experiment) is listed below as an example.

Page 1: Introduction

Welcome to our survey on crime.

This survey is expected to take approximately 10 minutes.

At the end of the survey, you will receive a unique, randomly generated code. Please copy and paste it on M-Turk in order to complete the HIT and receive your payment.

You must be 18 years old or older to participate in this survey.

Your participation is totally voluntary and you may quit the survey at any time.

☐ I am 18 years old or older. I have read and understood the information above and I want to participate in this research.

Page 2: Scenario Presentation

Please take the time to read the following important instructions.

Criminal law covers a very large number of different types of crimes. Some are considered to be very serious acts and others are not so serious. We are interested in your opinions about how serious you think different crimes are. We have made up descriptions of crimes. In this survey, you will be shown a description of one of these made-up crimes, asked a few questions to test your short-term memory of the details of that crime, then asked for your opinions about the crime.

In the next page, you will be presented with some text describing a crime scenario. Please read the scenario carefully. You will be asked to answer questions about what you read.

Page 3: Scenario

Please read the following scenario carefully. You will be asked questions about it in the following pages.

Tom Smith is a computer programmer who looks for security flaws on the Internet. He does this because he wants to \${e://Field/ConditionText}.

On June 3, 2013, Tom Smith found a security flaw in the Acme Insurance Company's website. He used that flaw to gain access to Acme's internal network and download 100,000 records from Acme's customer database. Each record consisted of a customer's full name, user ID, and password. Tom did not release the details of the flaw, and he did not use or release the records he downloaded. Acme's customers suffered no harm.

Pages 4.1–4.9: Memory Checks

Without going back to the previous page, please try to answer the following questions about the scenario you just read.

What were the consequences of Tom Smith's actions?

- ☐ He wants to make money by selling trade secrets
- ☐ He wants to learn about Internet security
- ☐ He wants to show how smart he is
- ☐ He wants to seek evidence of corporate corruption
- ☐ None of the above

If correct:

Correct! Tom wants to \${e://Field/Motivation}.

If incorrect:

Sorry, that is incorrect. Tom wants to \${e://Field/Motivation}.

Which one of the following kinds of data was among the data Tom Smith accessed?

- ☐ Credit card numbers
- ☐ Social security numbers
- ☐ Driver's license numbers
- ☐ Passwords Account numbers
- ☐ None of the above

If correct:

Correct! The data included **passwords**.

If incorrect:

Sorry, that is incorrect. The data included **passwords**.

What was Acme?

- ☐ Hospital
- ☐ Retailer
- ☐ Bank
- ☐ Manufacturer
- ☐ School
- ☐ Insurance company
- ☐ None of the above

If correct:

Correct! Acme was an **insurance company**.

If incorrect:

Sorry, that is incorrect. Acme was an **insurance company**.

How many records did Tom Smith download?

- ☐ 100
- ☐ 1,000
- ☐ 10,000
- ☐ 100,000
- ☐ 1,000,000
- ☐ None of the above

If correct:

Correct! Tom Smith downloaded **100,000** records.

If incorrect:

Sorry, that is incorrect. Tom Smith downloaded **100,000** records.

Page 5: Scenario Questions

Now please review again the scenario:

Tom Smith is a computer programmer who looks for security flaws on the Internet. He does this because he wants to \${e://Field/ConditionText}.

On June 3, 2013, Tom Smith found a security flaw in the Acme Insurance Company's website. He used that flaw to gain access to Acme's internal network and download 100,000 records from Acme's customer database. Each record consisted of a customer's full name, user ID, and password. Tom did not release the details of the flaw, and he did not use or release the records he downloaded. Acme's customers suffered no harm.

Please answer the following questions about this scenario:

How wrongful were Tom Smith's actions?						
Not at all wrongful						Extremely wrongful
1	2	3	4	5	6	7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How serious was the crime Tom Smith committed?						
Not at all serious						Extremely serious
1	2	3	4	5	6	7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How harshly should Tom Smith be punished?						
Not at all harshly						Extremely harshly
1	2	3	4	5	6	7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How responsible was the Acme Insurance Company for the crime?						
Not at all responsible						Extremely responsible
1	2	3	4	5	6	7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How clever was Tom Smith?						
Not at all clever						Extremely clever
1	2	3	4	5	6	7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How sensitive was the data that Tom Smith downloaded?						
Not at all sensitive						Extremely sensitive
1	2	3	4	5	6	7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Page 6: Specific Punishment

Please assume that Tom Smith is convicted in a court of law. How long of a jail or prison sentence do you believe would most appropriately fit the crime? Please assume that the entire term of a sentence would be served.

- ☐ No probation, jail, or prison time
- ☐ Probation only
- ☐ 0 to 29 days served in jail or prison
- ☐ 30 to 89 days served in jail or prison
- ☐ 90 to 179 days served in jail or prison
- ☐ 180 to 364 days served in jail or prison
- ☐ 1 year to less than 2 years served in jail or prison
- ☐ 2 years to less than 5 years served in jail or prison
- ☐ 5 years to less than 10 years served in jail or prison
- ☐ 10 years to less than 20 years served in jail or prison
- ☐ 20 or more years served in jail or prison
- ☐ Life served in jail or prison

Page 7: Potential Consequences

Please consider again the scenario:

Tom Smith is a computer programmer who looks for security flaws on the Internet. He does this because he wants to \${e://Field/ConditionText}.

On June 3, 2013, Tom Smith found a security flaw in the Acme Insurance Company's website. He used that flaw to gain access to Acme's internal network and download 100,000 records from Acme's customer database. Each record consisted of a customer's full name, user ID, and password. Tom did not release the details of the flaw, and he did not use or release the records he downloaded. Acme's customers suffered no harm.

Please think about what might have happened as a result of Tom Smith's actions.

How harmful might the *potential* consequences of Tom Smith's actions have been?

Not at all harmful							Extremely harmful	
1	2	3	4	5	6	7		
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

Page 8: CFIP

Here are some statements about personal information. From the standpoint of personal privacy, please indicate the extent to which you, as an individual, *agree* or *disagree* with each statement.

It usually bothers me when companies ask me for personal information.

Strongly disagree						Strongly agree	
1	2	3	4	5	6	7	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

All the personal information in computer databases should be double-checked for accuracy—no matter how much this costs.

Strongly disagree						Strongly agree	
1	2	3	4	5	6	7	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Companies should not use personal information for any purpose unless it has been authorized by the individuals who provided the information.

Strongly disagree						Strongly agree	
1	2	3	4	5	6	7	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Companies should devote more time and effort to preventing unauthorized access to personal information.

Strongly disagree						Strongly agree
1	2	3	4	5	6	7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

When companies ask me for personal information, I sometimes think twice before providing it.

Strongly disagree						Strongly agree
1	2	3	4	5	6	7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Companies should take more steps to make sure that the personal information in their files is accurate.

Strongly disagree						Strongly agree
1	2	3	4	5	6	7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

When people give personal information to a company for some reason, the company should never use the information for any other reason.

Strongly disagree						Strongly agree
1	2	3	4	5	6	7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Companies should have better procedures to correct errors in personal information.

Strongly disagree						Strongly agree
1	2	3	4	5	6	7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Computer databases that contain personal information should be protected from unauthorized access—no matter how much it costs.

Strongly disagree						Strongly agree
1	2	3	4	5	6	7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

It bothers me to give personal information to so many companies.

Strongly disagree						Strongly agree
1	2	3	4	5	6	7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Companies should never sell the personal information in their computer databases to other companies.

Strongly disagree						Strongly agree
1	2	3	4	5	6	7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Companies should devote more time and effort to verifying the accuracy of the personal information in their databases.

Strongly disagree						Strongly agree
1	2	3	4	5	6	7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Companies should never share personal information with other companies unless it has been authorized by the individuals who provided the information.

Strongly disagree						Strongly agree
1	2	3	4	5	6	7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Companies should take more steps to make sure that unauthorized people cannot access personal information in their computers.

Strongly disagree						Strongly agree
1	2	3	4	5	6	7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

I'm concerned that companies are collecting too much personal information about me.

Strongly disagree						Strongly agree
1	2	3	4	5	6	7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Page 9: Privacy Attitudes – Misc

How frequently have you personally been the victim of cybercrime or an invasion of privacy?

Never	Once	A few times	Several times
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

If you have been a victim of a cybercrime or invasion of privacy, can you please provide more details?

Some websites ask you to register with the site by providing personal information. When asked for such information, how often do you provide incorrect information?

- ☐ I have never given incorrect information
- ☐ Under 25% of the time
- ☐ 26%–50% of the time
- ☐ 51%–75% of the time
- ☐ 76% or more of the time

How much have you heard or read during the last year about the use and potential misuse of the information collected from the Internet?

- | | | | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Not at all | | | | | | Very much |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Page 10: Demographics

What is your gender?

- ☐ Male
- ☐ Female
- ☐ Prefer not to answer

What is your age?

- ☐ Under 18 years
- ☐ 18 to 24 years
- ☐ 25 to 34 years
- ☐ 35 to 44 years
- ☐ 45 to 54 years
- ☐ 55 to 64 years
- ☐ 65 years and over
- ☐ Prefer not to answer

In what country were you born?

[Drop-down list of countries]

In what country do you live now?

[Drop-down list of countries]

What is the highest level of education you have completed?

- ☐ Less than High School
- ☐ High School / GED
- ☐ Some College
- ☐ 2-year College Degree
- ☐ 4-year College Degree Masters Degree
- ☐ Professional Degree (JD, MD) Doctoral Degree
- ☐ Prefer not to answer

What is your current work situation?

[Drop-down list of work situations]

Which of the following most closely describes your current occupation?

[Drop-down list of occupations]

Page 11: Open-Ended

What do you think of Tom Smith?

What do you think the *actual* consequences of Tom Smith's actions were?

What do you think the *potential* consequences of Tom Smith's actions could have been?

Finally, what do you think this study was about?

Page 12: Conclusion

Thank you for taking part in our survey!

Please click on ">>" to complete your task and receive your unique, randomly generated compensation code. Remember to copy that code and paste it into the Mechanical Turk survey to be paid.

Please let us know if you have any comments about this study

Appendix E. Survey Text for the Chapter 3 Factorial Vignette Survey

Page 1: Introduction

Welcome to our survey on crime.

This survey is part of a research study conducted by Alessandro Acquisti at Carnegie Mellon University and Kirsten Martin of George Washington University. The purpose of the research is to understand people's opinions about crime.

Procedures

In this survey, you will be shown some descriptions of crimes and asked to rate the seriousness of those crimes. This survey is expected to take about 10-15 minutes to complete.

At the end of the survey, you will receive a unique, randomly generated code. Please copy and paste it on M-Turk in order to complete the HIT and receive your payment.

Participant Requirements

To participate in this study, you must be age 18 or older and a resident of the United States.

Risks

The risks and discomfort associated with participation in this study are no greater than those ordinarily encountered in daily life or during other online activities, and include boredom and fatigue.

Benefits

There may be no personal benefit from your participation in the study but the knowledge received may be of value to humanity.

Compensation & Costs

You will be paid \$1.75 for participating in this study. There will be no cost to you if you participate in this study.

Confidentiality

By participating in this study, you understand and agree that the data and information gathered during this study may be used by Carnegie Mellon for academic publications. You also understand and agree that Carnegie Mellon may be required to disclose your survey responses and other information as required by law, regulation, subpoena or court order. That information may include your IP address. We also collect information about your operating system and web browser.

Right to Ask Questions & Contact Information

If you have any questions about this study, feel free to ask them by contacting the Principal Investigator, Professor Alessandro Acquisti, Heinz College, Carnegie Mellon University, 5000 Forbes Av, Pittsburgh, PA 15213, acquisti@andrew.cmu.edu. If you have questions later, desire additional information, or wish to withdraw your participation, please contact the Principal Investigator by mail or e-mail in accordance with the contact information listed above.

If you have questions pertaining to your rights as a research participant; or to report objections to this study, you should contact the Research Regulatory Compliance Office at Carnegie Mellon University. Email: irb-review@andrew.cmu.edu. Phone: 412-268-1901 or 412-268-5460.

Voluntary Participation

Your participation in this research is voluntary and you may quit the survey at any time.

☐ I am 18 years old or older. I have read and understood the information above and I want to participate in this research.

Page 2: Instructions

Please take the time to read the following important instructions.

Criminal law covers a very large number of different types of crimes. Some are considered to be very serious acts and others are not so serious. We are interested in your opinions about how serious you think different crimes are.

In this survey, you will see twenty-five vignettes describing a made-up crime and its (also made-up) punishment. The details of the crime and the punishment given for the crime will change with each vignette.

Each vignette is a variation on the following scenario. Please read it carefully:

Tom Smith is a computer programmer who looks for security flaws on the Internet. On September 3, 2014, Tom found a security flaw in the website of an organization named ACR and used that flaw to download records from ACR's customer database. He anonymously released details about the flaw to the Internet, but did not use or release the records he downloaded. Before he did this, Tom had never been arrested or convicted of any crime.

ACR was \${e://Field/OrgTypeText0}.

Tom downloaded \${e://Field/RecordsText0} customer records.

Each record consisted of a customer's \${e://Field/DataText0}.

Tom's motivation was to \${e://Field/MotiveText0}.

ACR spent \${e://Field/OrgLossText0} to repair and secure its servers.

Its customers spent \${e://Field/CustLossText0} each to protect themselves from identity fraud.

Tom was convicted of the crime and received a sentence of \${e://Field/SentenceText0}.

The opening paragraph will be the same for each vignette, but the details following each vignette will change.

After each vignette, you will be asked for your personal opinion on how well the punishment fit the crime. A slider like the one below follows each vignette.



After each vignette, please use the slider to indicate whether you think the punishment was too low or too high and by how much. If you think the punishment was much too low, slide the marker to all the way to the right. If you think the punishment was much too high, move the marker all the way to the left. If you think the punishment falls somewhere between the two extremes, slide the marker to a position on the line that best shows how appropriate you think the punishment was.

For this example only, please move the marker on the slider above to a position about half way toward the left end of the line from its current position at the center of the line.

Please consider the following scenario (*N*/25):

Tom Smith is a computer programmer who looks for security flaws on the Internet. On September 3, 2014, Tom found a security flaw in the website of an organization named ACR and used that flaw to download records from ACR's customer database. He anonymously released details about the flaw to the Internet, but did not use or release the records he downloaded. Before he did this, Tom had never been arrested or convicted of any crime.

ACR was $\{e://Field/OrgTypeTextN\}$.

Tom downloaded $\{e://Field/RecordsTextN\}$ customer records.

Each record consisted of a customer's $\{e://Field/DataTextN\}$.

Tom's motivation was to $\{e://Field/MotiveTextN\}$.

ACR spent $\{e://Field/OrgLossTextN\}$ to repair and secure its servers.

Its customers spent $\{e://Field/CustLossTextN\}$ each to protect themselves from identity fraud. Tom was convicted of the crime and received a sentence of $\{e://Field/SentenceTextN\}$.



Note: This page was repeated for each of the 25 vignettes, with randomly generated values for each field on each page. The values for each field were:

- *OrgType*: “a bank,” “a non-profit organization,” “an insurance company,” “a government agency”
- *Records*: 10, 100, 1,000, 10,000, or 100,000
- *Data*: “e-mail address,” “full name, phone number, and address,” “full name, address, and social security number,” “full name, health history, medical diagnoses, and prescription records,” “full name, phone number, address, date of birth, and social security number,” “full name, user ID, and password”
- *Motive*: “learn about Internet security,” “seek evidence of corporate corruption,” “make money”
- *OrgLoss*: \$1000, \$10,000, \$100,000, \$1,000,000, \$10,000,000
- *CustLoss*: \$10, \$50, \$100, \$250, \$500
- *Sentence*: 3 months, 6 months, 1 year, 2 years, 5 years; plus either “probation,” “in jail” (for sentences less than 1 year) or “in prison” (for sentences of one year or more)

Thank you! We would now like to ask you some questions about yourself.

Here are some statements about personal information. From the standpoint of personal privacy, please indicate the extent to which you, as an individual, *agree* or *disagree* with each statement.

It usually bothers me when companies ask me for personal information.

Strongly disagree							Strongly agree
1	2	3	4	5	6	7	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

All the personal information in computer databases should be double-checked for accuracy—no matter how much this costs.

Strongly disagree							Strongly agree
1	2	3	4	5	6	7	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Companies should not use personal information for any purpose unless it has been authorized by the individuals who provided the information.

Strongly disagree							Strongly agree
1	2	3	4	5	6	7	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Companies should devote more time and effort to preventing unauthorized access to personal information.

Strongly disagree							Strongly agree
1	2	3	4	5	6	7	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

When companies ask me for personal information, I sometimes think twice before providing it.

Strongly disagree							Strongly agree
1	2	3	4	5	6	7	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Companies should take more steps to make sure that the personal information in their files is accurate.

Strongly disagree							Strongly agree
1	2	3	4	5	6	7	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

When people give personal information to a company for some reason, the company should never use the information for any other reason.

Strongly disagree							Strongly agree
1	2	3	4	5	6	7	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Companies should have better procedures to correct errors in personal information.

Strongly disagree							Strongly agree
1	2	3	4	5	6	7	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Computer databases that contain personal information should be protected from unauthorized access—no matter how much it costs.

Strongly disagree							Strongly agree
1	2	3	4	5	6	7	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

It bothers me to give personal information to so many companies.

Strongly disagree							Strongly agree
1	2	3	4	5	6	7	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Companies should never sell the personal information in their computer databases to other companies.

Strongly disagree							Strongly agree
1	2	3	4	5	6	7	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Companies should devote more time and effort to verifying the accuracy of the personal information in their databases.

Strongly disagree							Strongly agree
1	2	3	4	5	6	7	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Companies should never share personal information with other companies unless it has been authorized by the individuals who provided the information.

Strongly disagree							Strongly agree
1	2	3	4	5	6	7	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Companies should take more steps to make sure that unauthorized people cannot access personal information in their computers.

Strongly disagree							Strongly agree
1	2	3	4	5	6	7	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

I'm concerned that companies are collecting too much personal information about me.

Strongly
disagree

Strongly
agree

1

2

3

4

5

6

7

☐

☐

☐

☐

☐

☐

☐

Page 30: Privacy Attitudes – Misc

How frequently have you personally been the victim of cybercrime or an invasion of privacy?

Never

Once

A few times

Several times

☐

☐

☐

☐

If you have been a victim of a cybercrime or invasion of privacy, can you please provide more details?

Some websites ask you to register with the site by providing personal information. When asked for such information, how often do you provide incorrect information?

- ☐ I have never given incorrect information
- ☐ Under 25% of the time
- ☐ 26%–50% of the time
- ☐ 51%–75% of the time
- ☐ 76% or more of the time

How much have you heard or read during the last year about the use and potential misuse of the information collected from the Internet?

Not at all

Very much

1

2

3

4

5

6

7

☐

☐

☐

☐

☐

☐

☐

Page 31: Demographics

What is your gender?

- ☐ Male
- ☐ Female
- ☐ Prefer not to answer

What is your age?

- ☐ Under 18 years
- ☐ 18 to 24 years
- ☐ 25 to 34 years
- ☐ 35 to 44 years
- ☐ 45 to 54 years
- ☐ 55 to 64 years
- ☐ 65 years and over
- ☐ Prefer not to answer

In what country were you born?

[Drop-down list of countries]

In what country do you live now?

[Drop-down list of countries]

What is the highest level of education you have completed?

- ☐ Less than High School
- ☐ High School / GED
- ☐ Some College
- ☐ 2-year College Degree
- ☐ 4-year College Degree Masters Degree
- ☐ Professional Degree (JD, MD) Doctoral Degree
- ☐ Prefer not to answer

What is your current work situation?

[Drop-down list of work situations]

Which of the following most closely describes your current occupation?

[Drop-down list of occupations]

Page 32: Open-Ended

Finally, what do you think this study was about?

Page 33: Conclusion

Thank you for taking part in our survey!

Please click on ">>" to complete your task and receive your unique, randomly generated compensation code. Remember to copy that code and paste it into the Mechanical Turk survey to be paid.

Please let us know if you have any comments about this study

Appendix F. Distribution of Ratings by Vignette in the Chapter 4 Perceptions Study

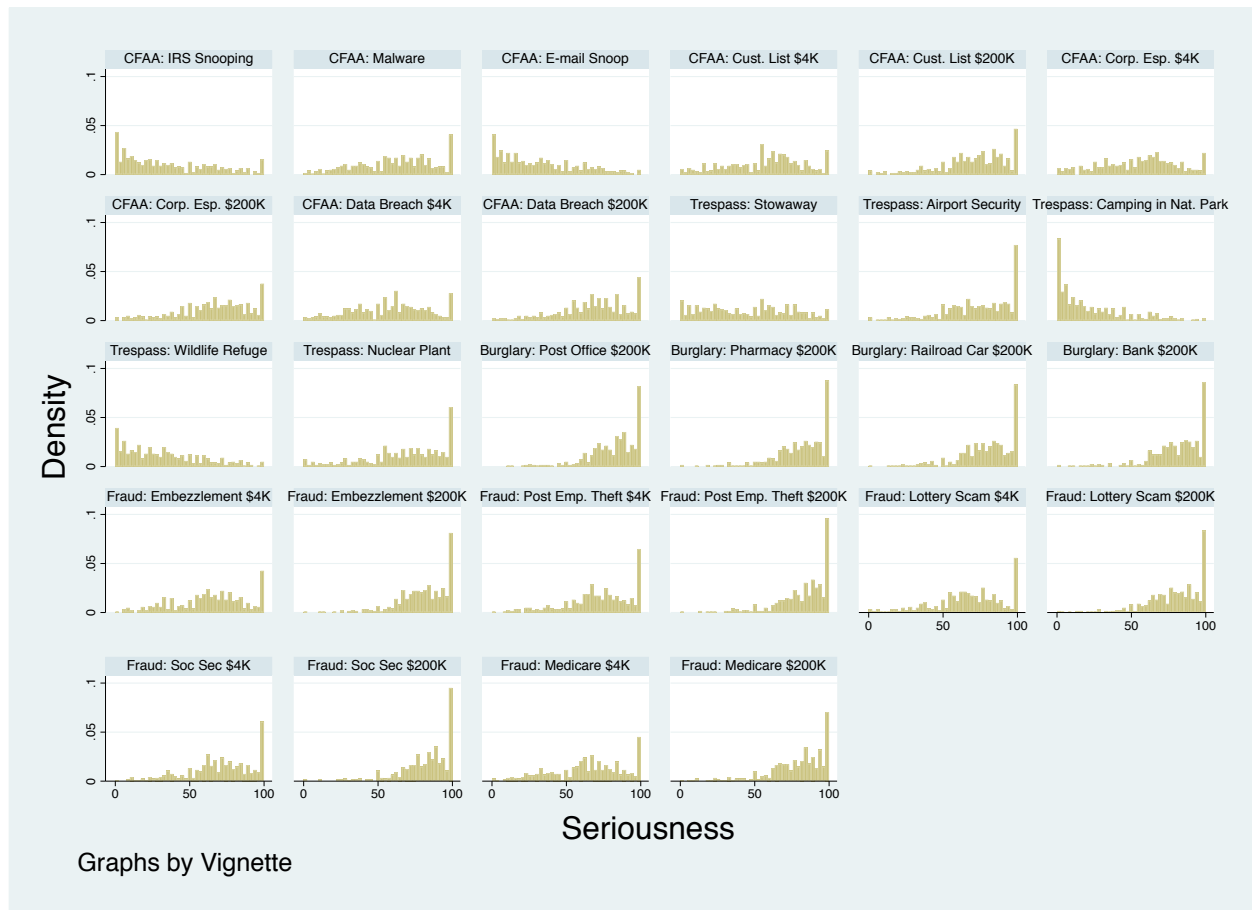


Figure 12: Distribution of ratings by vignette: Seriousness

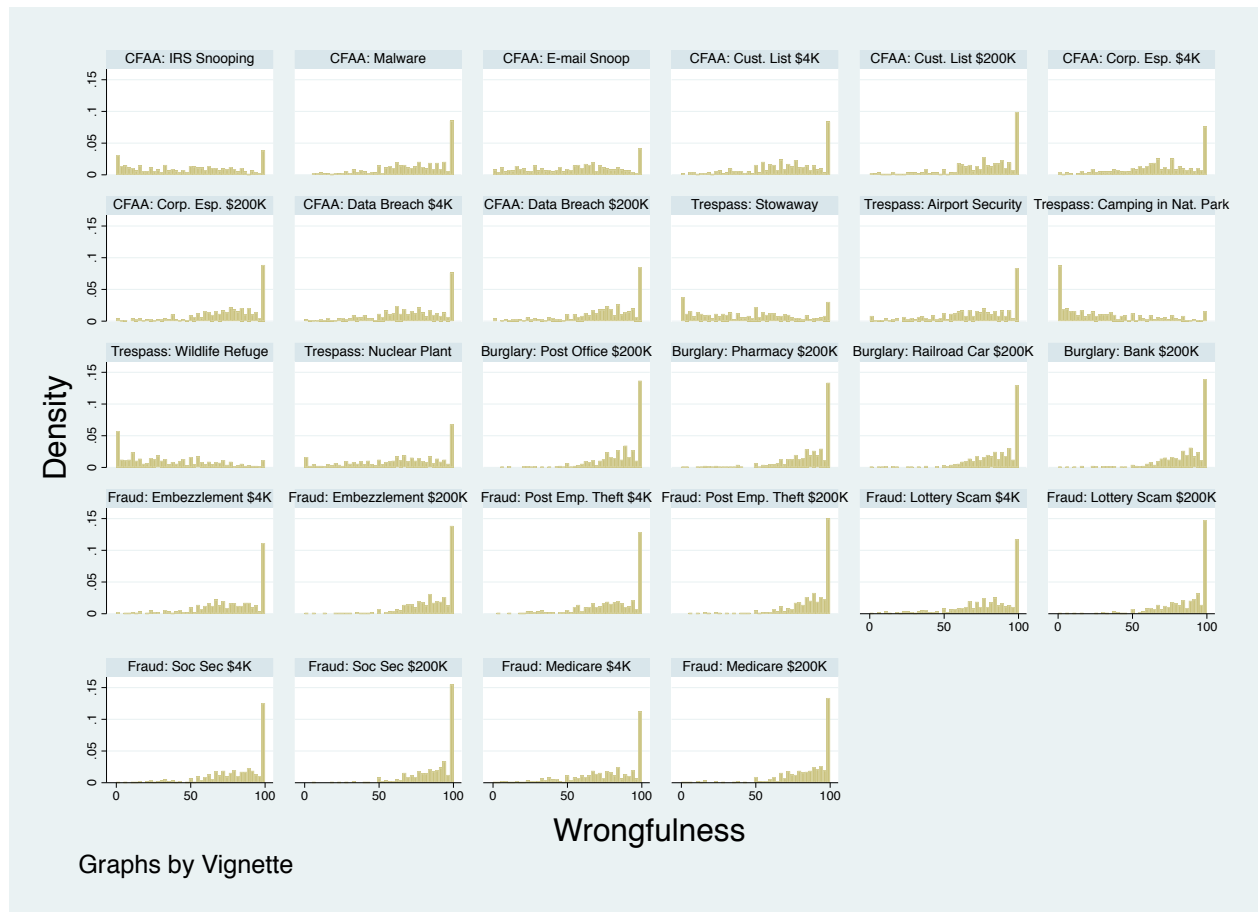


Figure 13: Distribution of ratings by vignette: Wrongfulness

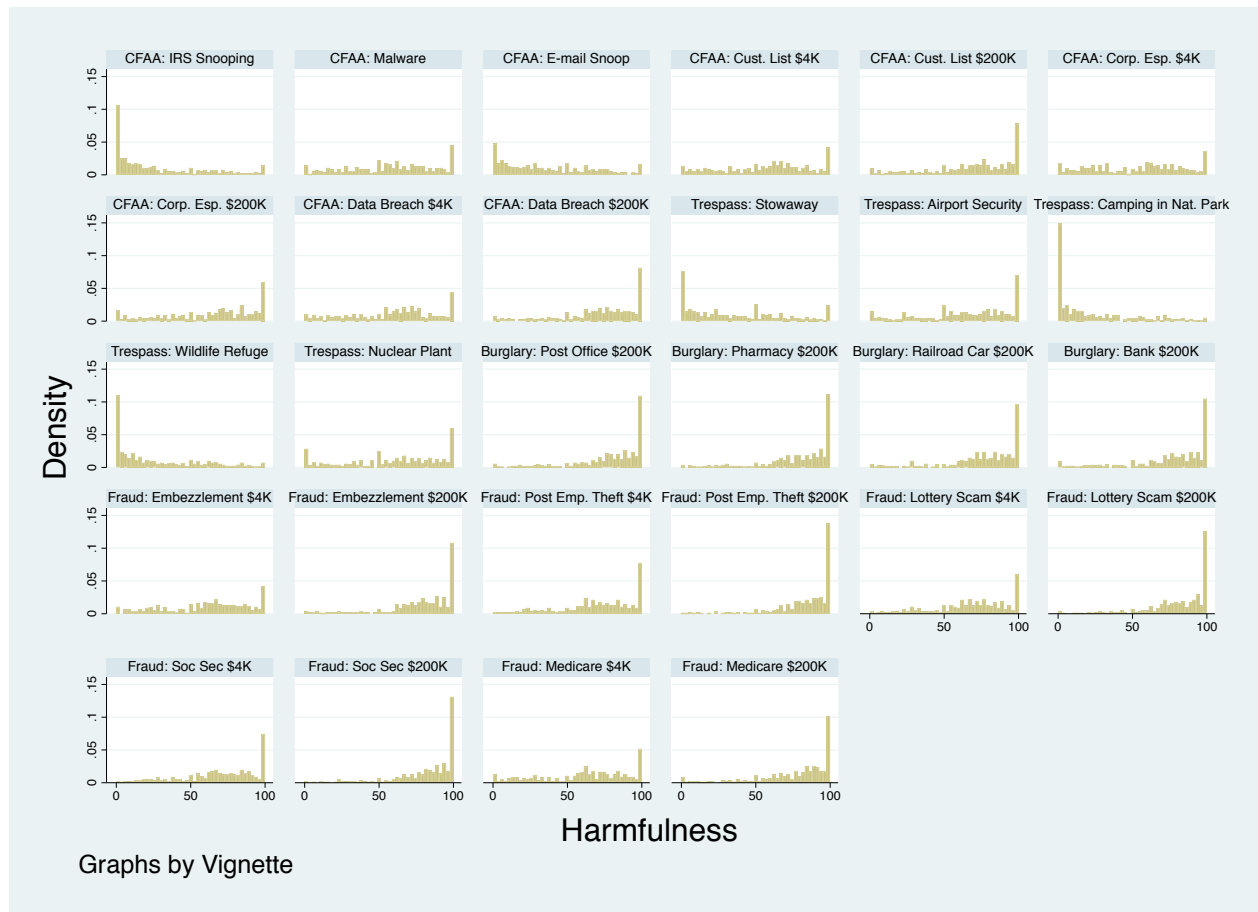


Figure 14: Distribution of ratings by vignette: Harmfulness

Appendix G. Data Quality Checks for the Chapter 4 Perceptions Study

G.1. Inconsistently Rated Vignette Pairs

The vignette set included eight pairs of vignettes that had the same crime description but different amounts of loss (either \$4,000 or \$200,000). The vignette with the higher dollar amount thus loosely dominates the vignette with the lower amount—that is, an attentive and careful participant should not rate a crime that led to \$4,000 in losses as being more serious, wrongful, or harmful than the same crime with \$200,000 in losses (although they might reasonably rate the crimes as being the same; thus “loosely” dominated). These dominated vignette pairs (“DVPs”) therefore can serve as a form of attention or response-quality check. We could presume that respondents who rated several of these pairs inconsistently (i.e., with the \$4,000 crime as worse than the \$200,000 crime) for a particular rating type (serious, wrongful, or harmful) either were not paying attention, did not notice that the vignettes were the same, or, when ratings were close, did not use the rating talk slider with as much precision as I had hoped. A pair of vignettes for which a participant rated the \$4,000 version as worse than the \$200,000 version can be referred to as an inconsistently-rated dominated vignette pair, or “IRDVP.”

Overall, nearly 84% of respondents rated at least one of the 24 DVPs inconsistently. Of the 11,976 total DVPs across all ratings and responses, 2,122 (18%) were rated inconsistently. Figure 15 shows the distribution of IRDVPs per respondent.

To allow for imprecision in the rating task, the criteria for an IRDVP can be loosened to the \$4,000 version of a vignette being rated higher than the \$200,000 version by 5 points or more. About 56% of respondents had at least one IRDVP under those criteria, and 9% of all DVPs (1,047/11,976) were rated inconsistently. Figure 16 shows this distribution.

No.	Freq.	Percent	Cum.
0	68	13.63	13.63
1	74	14.83	28.46
2	76	15.23	43.69
3	48	9.62	53.31
4	42	8.42	61.72
5	28	5.61	67.33
6	32	6.41	73.75
7	23	4.61	78.36
8	30	6.01	84.37
9	25	5.01	89.38
10	15	3.01	92.38
11	13	2.61	94.99
12	10	2.00	96.99
13	6	1.20	98.20
14	4	0.80	99.00
15	3	0.60	99.60
16	1	0.20	99.80
18	1	0.20	100.00
Total	499	100.00	

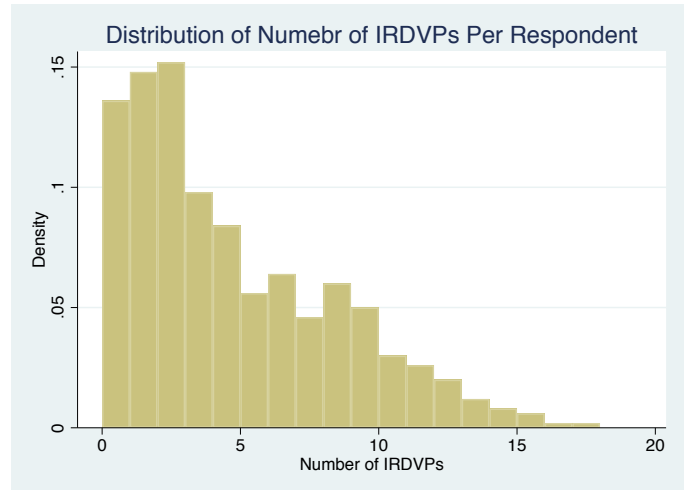


Figure 15: Distribution of IRDVPs per Respondent

No.	Freq.	Percent	Cum.
0	165	33.07	33.07
1	114	22.85	55.91
2	64	12.83	68.74
3	45	9.02	77.76
4	38	7.62	85.37
5	24	4.81	90.18
6	14	2.81	92.99
7	6	1.20	94.19
8	13	2.61	96.79
9	3	0.60	97.39
10	6	1.20	98.60
11	5	1.00	99.60
12	1	0.20	99.80
14	1	0.20	100.00
Total	499	100.00	

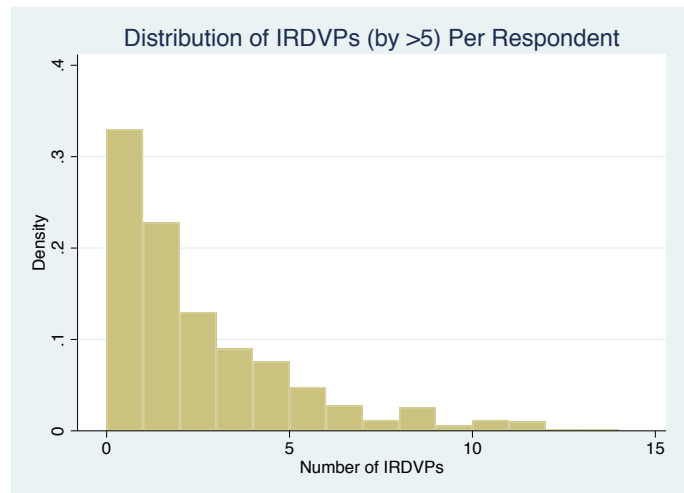


Figure 16: Distribution of IRDVPs where the difference in ratings is more than 5 points on the 100-point scale

G.1.1. Comparison to the Factorial Vignette Experiments

In the factorial vignette experiment (Section 3.4), 197 vignette pairs (out of 5575) were loosely dominated. Domination was rarer and in the factorial study given the number of factors for each vignette; only if *every* factor in vignette *A* was greater than or equal to *every* factor in vignette *B* would *A* loosely dominate *B* (because *A* has to be at least as bad as *B*). Participants rated 23 (11.7%) of those loosely dominated pairs inconsistently (i.e., as *B* being worse than *A*).

In the Chapter 4 survey, participants rated 2,122 of 11,976 DVPs inconsistently (17.7%). 1,047 (8.7%) were off by more than 5 points on the 100-point scale (8.7%).

G.1.2. Effect of the Space Between Vignettes in the Display Order on IRDVPs

Participants may have been more likely to have rated dominated vignette pairs inconsistently when the pairs were widely separated in the presentation of vignettes (e.g., if a \$4,000 version of a vignette appeared near the beginning and the \$200,000 version appears next to the end, as opposed to both vignettes appearing near each other). Figure 17 shows the effect of distance in the presentation order on IRDVPs by graphing, the percentage of all DVPs that were inconsistently rated according to distance between the members of the vignette pairs (e.g., an ordering distance of “1” means the vignettes were adjacent). Presentation distance does appear to have had some effect.

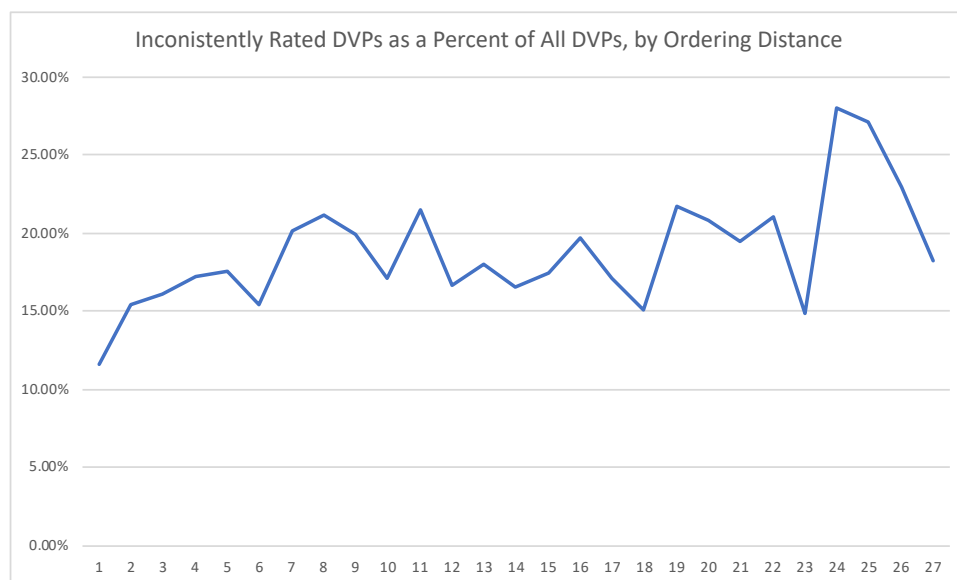


Figure 17: Inconsistently rated IRDVPs as a percent of all DVPs, by ordering distance

G.2. Overall Distribution of Ratings

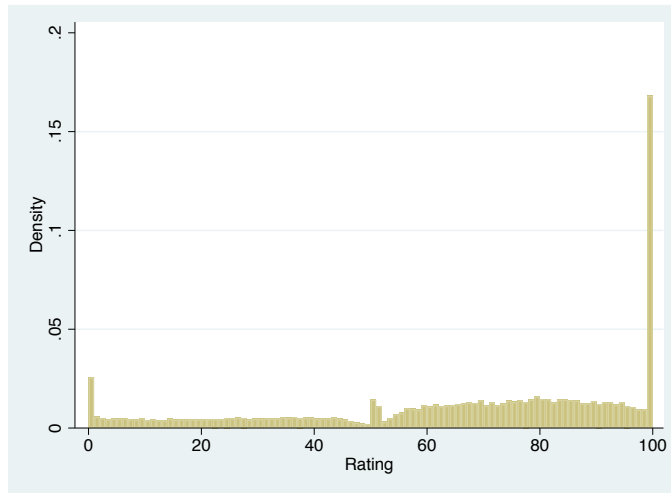


Figure 18: Overall distribution of ratings

Figure 18 shows the overall distribution of all ratings in the experiment. About 16% of all ratings were 100. The lack of clustering at the midpoint is notable given that the midpoint was the default position for the slider rating task.

Figure 19 shows the distribution of 100 ratings for all respondents. A few participants were more likely to rate most of their vignettes at the maximum. About a tenth of participants (54, 10.8%) rated at least half the vignettes at 100, and 15 (3%) rated three-quarters of their vignettes at 100. Three users (0.6%) rated *all* their vignettes at 100.

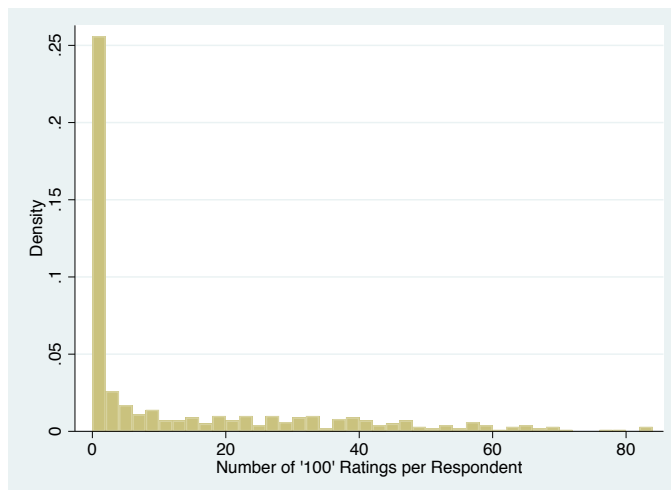
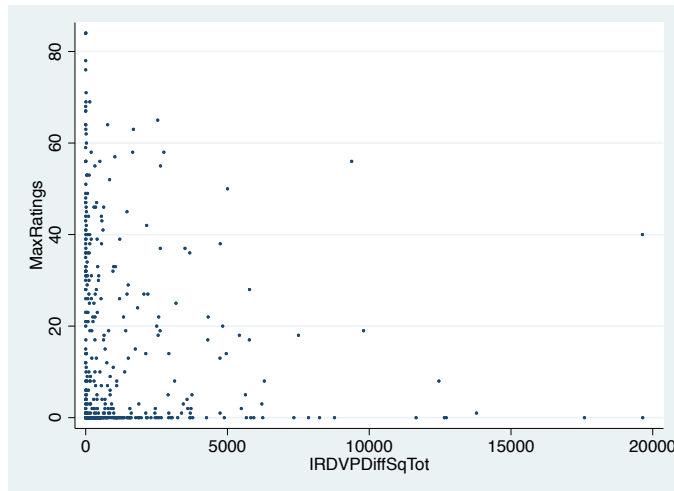


Figure 19: Distribution of “100” ratings per respondent

There was no apparent correlation between number of “100” ratings and sum of squares IRDVP distance:



G.3. Effect of Completion Time on Results

Participants who completed the survey quickly tended to have more IRDVPs and ratings of “100,” but the effect does not appear pronounced. Figure 20 shows the distribution of number IRDVPs by completion time. Figure 21 show the distribution of “100” ratings by completion time.

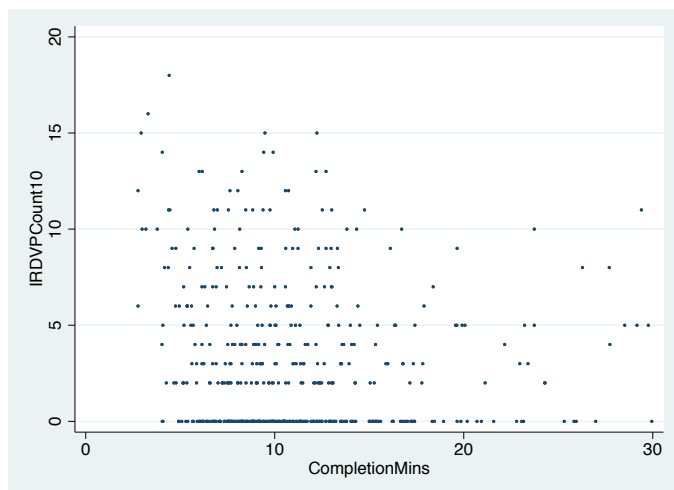


Figure 20: IRDVPs (by 10 or more) by how long the respondent took to answer the survey

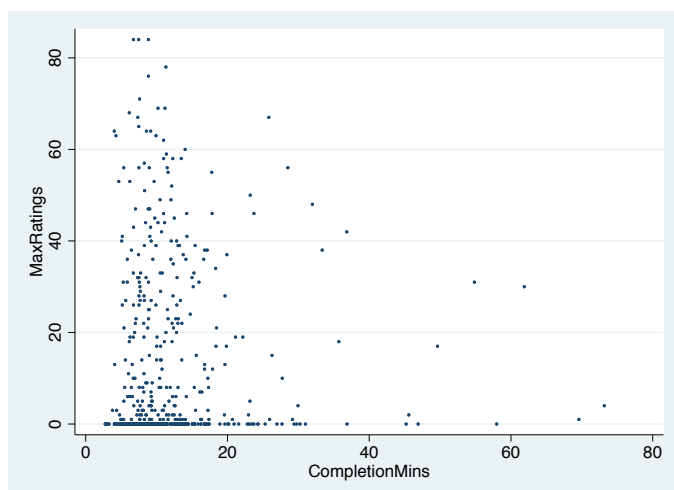


Figure 21: Number of “100” Ratings by Completion Time

G.4. Ordering Effects

G.4.1. Intra-Page Display Order

The order in which vignettes were displayed on a page had no meaningful effect on the mean or variance of ratings. In other words, the last vignette on each page—regardless of which crime it described on a given survey—received similar overall ratings to the first vignette displayed on each page. Figure 22 shows the distribution of ratings by display order.

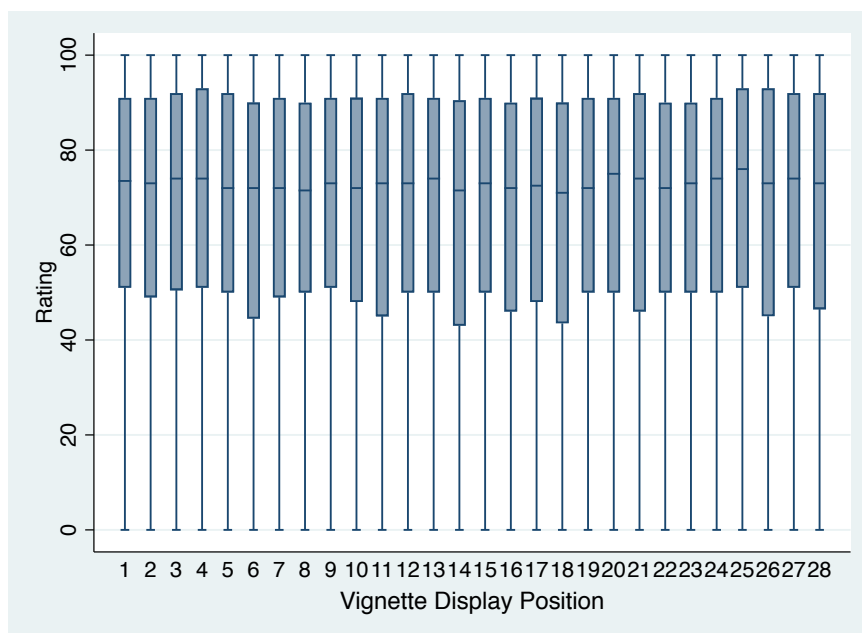


Figure 22: Distribution of Responses by Display Order

G.4.2. Page Order

There may have been some fatigue effect on the second and third survey pages. Recall that the first rating page in the survey asked for ratings of seriousness, the next page asked for ratings of wrongfulness on the same vignettes, and the third page asked for ratings of harmfulness. When reviewing responses, I noticed that some participants seemed to have reasonable answers for seriousness but nearly all 100s on the wrongfulness and harmfulness pages.

The fatigue effect is not monotonic, however. The total number of vignettes rated at 0 or 100 increased on the second page (wrongfulness) but then *decreased* on the third page (harmfulness):

Serious: 1,771
Wrongful: 3,211
Harmful: 2,652

The same pattern holds for the number of inconsistently rated dominated vignette pairs (IRDVPs), increasing on the second rating page but decreasing on the third (number of those IRDVPs that were by more than 5 points in parentheses):

Serious: 643 (318)
Wrongful: 782 (377)
Harmful: 697 (352)

The number of people with at least one IRDVP (within 5 points, to allow for rating task fuzziness) for that rating type also increased slightly on the second page but decreased on the third:

Serious: 190 (38%)
Wrongful: 207 (41%)
Harmful: 198 (40%)

These results are puzzling. Had the page order been Serious → Harmful → Wrongful (instead of the actual order of Serious → Wrongful → Harmful), the pattern might have indicated that fatigue effects increased with each page, as might be expected. I checked the survey page in Qualtrics, the data set, and my Stata code to verify that the ratings for wrongfulness and harmfulness were not swapped. The difference in 100 ratings might be explained by participant beliefs that crimes are extremely wrong, but I cannot explain why IRDVPs would increase on the second page then decrease on the third.

G.5. Correlation Between Rating Types

There is a strong correlation in the overall data set between the rated seriousness, wrongfulness, and harmfulness of a given crime vignette: if people find a crime to be very serious, they are also likely to find it more harmful and wrongful. The correlation is not necessarily universal, of course—someone could believe that an IRS worker accessing tax returns without a valid purpose is very wrong but not especially harmful. If so, however, the overall seriousness will probably reflect that belief.

This relationship allows another method of measuring the quality of a response. A participant who is paying attention to the vignettes is likely to have a statistically significant correlation between ratings of seriousness, wrongfulness, and harmfulness. Someone who is not paying attention almost certainly will not.

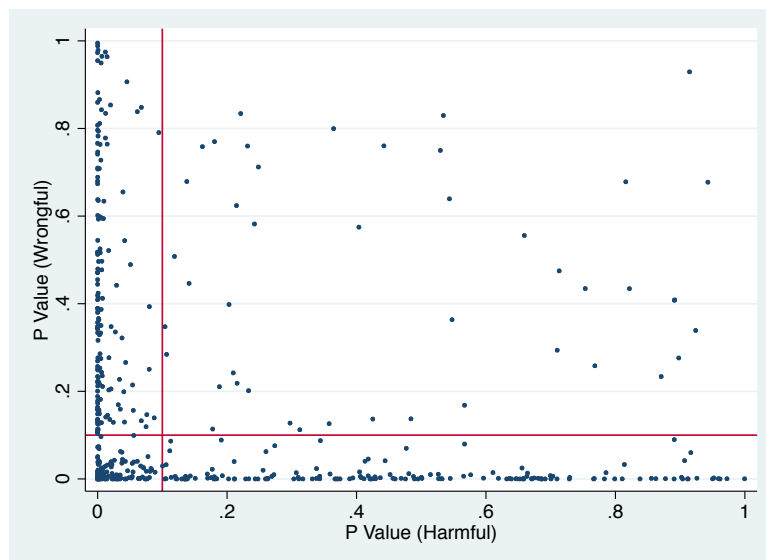


Figure 23: *P* values of individual regressions on ratings types (DV: Seriousness)

Figure 23 is a scatter plot of the *p*-values resulting from individual regressions of the form $Seriousness = \beta_1 Harmfulness + \beta_2 Wrongfulness + c$.

Of 499 total responses, 450 (90%) had $p < 0.10$ for either wrongfulness or harmfulness on individual regressions. There were also 6 responses that had no coefficients (and therefore no *p*-values) because of co-linearity in the model. For each respondent, the co-linearity was a result of rating all the vignettes as 0 or 100.

The R^2 value resulting from individual regressions could indicate the strength of correlation between ratings of seriousness, wrongfulness, and harmfulness. Figure 24 shows the distribution of R^2 values in the sample. But R^2 values do not seem to be useful for excluding “unreliable” answers, because of the inherent arbitrariness in setting acceptable vs. unacceptable R^2 values.

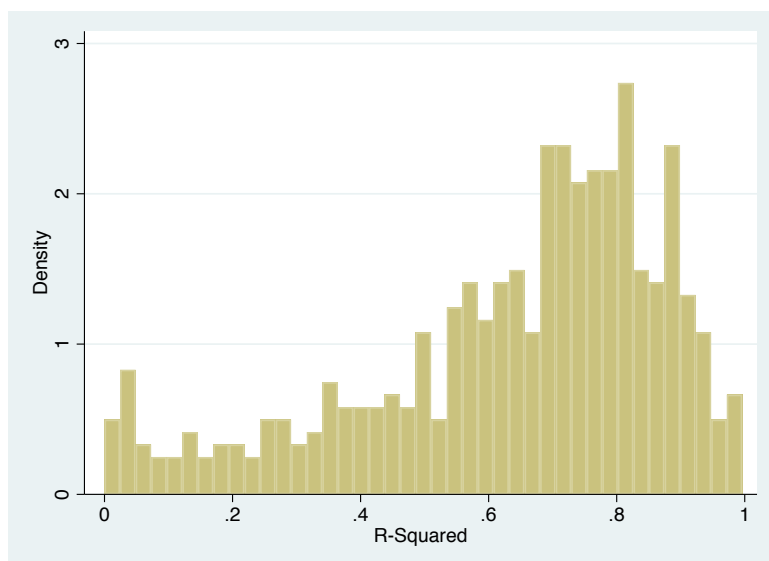


Figure 24: Distribution of R^2 values from individual regressions on rating types

G.6. Regression Results When Excluding “Unreliable” Responses

The next three tables list regression results when unreliable answers are excluded. Three levels of exclusion are used: “loose,” “medium,” and “strict.”

“Loose” exclusion criteria exclude from the regression and answers with:

- more than one IRDVP of that rating type by more than 5 points (i.e., for the regression on ratings of Seriousness, only the vignette pairs displayed for that rating page count, not IRDVPs for the other two pages),
- 20 or more ratings of 0 or 100 (out of 28 possible) on that page, or
- p -values that are ≥ 0.1 for both wrongfulness and harmfulness in the individual regressions on seriousness (i.e., the correlation between seriousness and both wrongfulness and harmfulness are insignificant at $p < 0.1$).

The “medium” exclusion criteria exclude answers with:

- *any* IRDVPs of that rating page by 5 points or more (although the respondent might have other IRDVPs on the other rating pages),
- more than half the ratings on a page at the extremes (i.e., more than 14), or
- p -values that are ≥ 0.1 for both wrongfulness and harmfulness in the individual regressions on seriousness.

The “strict” criteria is the same as the “medium” criteria except that it excludes answers from respondents who had a sum of squared IRDVP differences of more than 10 over *all* rating types. The strict exclusion criteria thus includes in the analysis responses with one IRDVP by 3 or fewer points, two by 2 points each, one by 1 point and another by 3, 10 by 1 point each, etc.

Table 38 summarizes the exclusion criteria at each level, and Table 39 shows the regression results when these exclusion criteria are applied.

Table 38: Summary of data-quality exclusion criteria

	Loose	Medium	Strict
IRDVPs	More than 1 (by 5+) (for that rating type only)	Any (by 5+) (for that rating type only)	Sum of squared IRDVP differences (over all rating types) is ≥ 10
Ratings of 0 or 100	≥ 20 on that rating type	≥ 14 on that rating type	≥ 14 on that rating type
P-values in indiv. regressions of wrongfulness and harmfulness on seriousness	Neither $P < 0.1$	Neither $P < 0.1$	Neither $P < 0.1$

Table 39: Regressions with exclusions

	Loose			Medium			Strict		
	Serious	Wrongful	Harmful	Serious	Wrongful	Harmful	Serious	Wrongful	Harmful
Crime Type (v. CFAA)									
Trespass	-17.010*** (1.044)	-24.794*** (1.277)	-16.527*** (1.196)	-15.531*** (1.106)	-24.553*** (1.486)	-16.743*** (1.383)	-16.879*** (2.039)	-23.698*** (2.188)	-15.958*** (1.893)
Burglary	41.312*** (1.005)	30.822*** (1.247)	43.022*** (1.381)	42.083*** (1.203)	32.194*** (1.490)	43.122*** (1.699)	44.908*** (1.931)	37.726*** (2.152)	47.843*** (2.339)
Fraud	14.842*** (0.744)	12.946*** (0.911)	12.898*** (0.989)	14.223*** (0.885)	12.356*** (1.062)	10.478*** (0.996)	13.820*** (1.711)	11.647*** (1.800)	10.450*** (1.793)
Offense Level (vs. 4)									
6	12.965*** (0.883)	9.904*** (1.194)	15.948*** (1.124)	12.337*** (1.070)	9.369*** (1.413)	15.260*** (1.330)	10.146*** (1.956)	10.058*** (1.911)	13.542*** (2.032)
8	12.867*** (1.101)	9.773*** (1.357)	17.274*** (1.339)	11.741*** (1.259)	8.144*** (1.676)	17.662*** (1.572)	10.519*** (2.126)	8.807*** (2.253)	16.478*** (2.378)
16	26.936*** (0.999)	19.564*** (1.305)	32.091*** (1.321)	28.546*** (1.207)	21.363*** (1.569)	35.069*** (1.528)	30.755*** (2.245)	26.097*** (2.340)	39.036*** (2.398)
18	27.113*** (1.123)	20.190*** (1.372)	36.185*** (1.349)	29.181*** (1.328)	22.172*** (1.644)	39.217*** (1.516)	32.174*** (2.377)	27.792*** (2.349)	44.183*** (2.306)
Crime Type x Off. Level									
Trespass x 6	5.516*** (1.605)	1.469 (1.839)	-3.820* (1.761)	5.097** (1.924)	3.016 (2.199)	-3.670 (2.059)	7.136* (3.503)	6.080 (3.629)	-2.030 (3.115)
Trespass x 8	33.197*** (1.571)	25.975*** (2.039)	22.630*** (1.894)	33.750*** (1.780)	28.052*** (2.574)	21.259*** (2.405)	37.745*** (3.071)	29.476*** (3.433)	23.691*** (3.361)
Fraud x 6	-3.693*** (0.639)	-2.212*** (0.664)	-4.357*** (0.714)	-4.018*** (0.673)	-3.301*** (0.746)	-3.013*** (0.791)	-3.716** (1.156)	-3.749*** (1.008)	-1.904 (1.359)
Demog. Ctrl.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	49.102*** (4.885)	60.224*** (5.131)	44.391*** (5.996)	53.555*** (7.690)	57.336*** (6.499)	41.171*** (6.372)	65.178*** (9.562)	59.050*** (12.489)	61.799*** (14.412)
N	10472	8288	10052	7252	4984	6384	2436	1932	2324
Clusters	374	296	359	259	178	228	87	69	83
R2 (Within)	0.528	0.493	0.479	0.552	0.534	0.523	0.604	0.595	0.586
R2 (Between)	0.086	0.105	0.101	0.085	0.145	0.138	0.318	0.281	0.406
R2 (Overall)	0.423	0.401	0.387	0.453	0.443	0.432	0.543	0.529	0.545

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The regression results are remarkably robust to exclusion of questionable responses. Even with the strict exclusions that eliminate between 83% and 86% of responses from analysis, the main coefficients remain statistically significant (although some of the interactions lose significance). The coefficients also stay fairly consistent as the exclusions become more strict.

G.7. Regressions by Individual Vignettes

Table 40: Regressions by individual vignette, with “loose” exclusions

	Serious		Wrongful		Harmful	
Vignette (Baseline: CFAA: E-mail Snooping)						
CFAA: IRS Snooping	3.278*	(1.370)	-6.679***	(1.495)	-9.178***	(1.333)
CFAA: Malware	30.313***	(1.222)	19.588***	(1.354)	23.780***	(1.378)
CFAA: Cust. List \$4K	24.944***	(1.237)	15.416***	(1.602)	21.008***	(1.506)
CFAA: Cust. List \$200K	40.690***	(1.384)	25.861***	(1.665)	38.813***	(1.699)
CFAA: Corp. Esp. \$4K	22.273***	(1.272)	12.368***	(1.647)	17.563***	(1.474)
CFAA: Corp. Esp. \$200K	34.957***	(1.365)	21.527***	(1.680)	31.192***	(1.714)
CFAA: Data Breach \$4K	25.270***	(1.278)	14.838***	(1.656)	23.875***	(1.518)
CFAA: Data Breach \$200K	38.751***	(1.353)	24.213***	(1.807)	40.869***	(1.626)
Trespass: Stowaway	12.668***	(1.523)	-9.118***	(1.769)	0.468	(1.628)
Trespass: Airport Security	43.286***	(1.439)	18.014***	(1.821)	30.499***	(1.786)
Trespass: Camping in Nat. Park	-10.872***	(1.217)	-23.402***	(1.613)	-14.031***	(1.414)
Trespass: Wildlife Refuge	-0.754	(1.371)	-17.581***	(1.552)	-9.290***	(1.493)
Trespass: Nuclear Plant	37.217***	(1.630)	12.500***	(1.816)	25.989***	(1.838)
Burglary: Post Office \$200K	53.128***	(1.269)	35.764***	(1.561)	48.560***	(1.715)
Burglary: Pharmacy \$200K	53.265***	(1.298)	35.071***	(1.671)	49.114***	(1.671)
Burglary: Railroad Car \$200K	50.623***	(1.260)	34.632***	(1.563)	47.114***	(1.624)
Burglary: Bank \$200K	53.021***	(1.274)	35.034***	(1.628)	46.769***	(1.730)
Fraud: Embezzlement \$4K	32.773***	(1.254)	23.176***	(1.608)	26.192***	(1.583)
Fraud: Embezzlement \$200K	52.821***	(1.240)	36.145***	(1.527)	48.682***	(1.652)
Fraud: Post Emp. Theft \$4K	41.385***	(1.290)	27.936***	(1.569)	35.972***	(1.524)
Fraud: Post Emp. Theft \$200K	55.286***	(1.277)	38.628***	(1.514)	55.284***	(1.467)
Fraud: Lottery Scam \$4K	36.428***	(1.270)	26.108***	(1.588)	34.106***	(1.456)
Fraud: Lottery Scam \$200K	51.019***	(1.315)	36.250***	(1.572)	52.616***	(1.524)
Fraud: Soc Sec \$4K	39.770***	(1.285)	27.632***	(1.620)	34.604***	(1.494)
Fraud: Soc Sec \$200K	54.642***	(1.229)	37.949***	(1.556)	52.830***	(1.480)
Fraud: Medicare \$4K	33.393***	(1.269)	24.017***	(1.610)	27.273***	(1.535)
Fraud: Medicare \$200K	51.463***	(1.212)	36.345***	(1.573)	48.056***	(1.584)
Demographic Controls	Yes		Yes		Yes	
Constant	37.905***	(4.899)	55.921***	(5.184)	39.524***	(5.951)
N	10472		8288		10052	
Clusters	374		296		359	
R2 (Within)	0.571		0.522		0.513	
R2 (Between)	0.086		0.105		0.101	
R2 (Overall)	0.455		0.422		0.413	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 41: Regressions by individual vignette, with “strict” exclusions

	Serious		Wrongful		Harmful	
Vignette (Baseline: CFAA: E-mail Snooping						
CFAA: IRS Snooping	5.333*	(2.155)	-9.725***	(2.953)	-10.060***	(2.556)
CFAA: Malware	31.149***	(2.430)	17.667***	(2.635)	20.361***	(2.466)
CFAA: Cust. List \$4K	23.414***	(2.611)	14.087***	(2.667)	17.458***	(2.642)
CFAA: Cust. List \$200K	45.655***	(2.869)	30.319***	(2.999)	43.241***	(3.132)
CFAA: Corp. Esp. \$4K	20.920***	(2.518)	11.362***	(2.824)	14.325***	(2.689)
CFAA: Corp. Esp. \$200K	40.368***	(2.923)	26.942***	(2.977)	38.410***	(3.085)
CFAA: Data Breach \$4K	22.586***	(2.730)	12.667***	(3.155)	19.145***	(2.683)
CFAA: Data Breach \$200K	42.724***	(2.905)	28.971***	(3.289)	45.759***	(2.805)
Trespass: Stowaway	12.563***	(3.277)	-4.913	(3.698)	-1.012	(3.299)
Trespass: Airport Security	45.989***	(2.795)	18.493***	(3.865)	29.771***	(3.301)
Trespass: Camping in Nat. Park	-8.230***	(2.457)	-22.203***	(2.976)	-14.024***	(2.345)
Trespass: Wildlife Refuge	-1.207	(2.506)	-19.899***	(2.604)	-11.024***	(2.613)
Trespass: Nuclear Plant	41.103***	(3.467)	15.971***	(3.759)	25.518***	(3.335)
Burglary: Post Office \$200K	57.621***	(2.460)	40.087***	(2.743)	51.133***	(3.263)
Burglary: Pharmacy \$200K	57.460***	(2.462)	40.623***	(2.804)	53.928***	(2.656)
Burglary: Railroad Car \$200K	56.080***	(2.401)	40.203***	(2.716)	51.036***	(2.808)
Burglary: Bank \$200K	57.115***	(2.450)	40.580***	(2.811)	49.012***	(3.192)
Fraud: Embezzlement \$4K	30.989***	(2.409)	20.261***	(2.623)	22.687***	(3.140)
Fraud: Embezzlement \$200K	57.667***	(2.468)	41.043***	(2.734)	52.735***	(2.964)
Fraud: Post Emp. Theft \$4K	39.678***	(2.605)	23.087***	(2.815)	30.566***	(2.915)
Fraud: Post Emp. Theft \$200K	60.092***	(2.518)	42.609***	(2.655)	58.386***	(2.584)
Fraud: Lottery Scam \$4K	33.322***	(2.331)	23.116***	(3.052)	30.157***	(2.707)
Fraud: Lottery Scam \$200K	56.218***	(2.573)	41.565***	(2.825)	57.747***	(2.675)
Fraud: Soc Sec \$4K	35.138***	(2.363)	21.812***	(2.938)	29.108***	(2.760)
Fraud: Soc Sec \$200K	57.414***	(2.445)	40.319***	(2.772)	54.325***	(2.605)
Fraud: Medicare \$4K	31.103***	(2.413)	19.739***	(2.865)	24.771***	(2.837)
Fraud: Medicare \$200K	55.126***	(2.390)	39.812***	(2.790)	51.699***	(2.791)
Demographic Controls	Yes		Yes		Yes	
Constant	53.017***	(9.557)	56.402***	(12.366)	58.365***	(14.352)
N	2436		1932		2324	
Clusters	87		69		83	
R2 (Within)	0.642		0.621		0.613	
R2 (Between)	0.318		0.281		0.406	
R2 (Overall)	0.573		0.550		0.565	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix H. Survey Text for the Chapter 4 Perceptions Study

Page 1: Introduction

Welcome to our survey on crime.

This survey is part of a research study conducted by Alessandro Acquisti and Jim Graves at Carnegie Mellon University. The purpose of the research is to understand people's opinions about crime.

Study Title: STUDY2018_00000301

Principal Investigator:

Jim Graves

Engineering and Public Policy 129 Baker Hall

5000 Forbes Avenue Pittsburgh, PA 15213 412-268-2670

jtg@cmu.edu

Faculty Advisor: Dr. Alessandro Acquisti

Procedures

In this survey, you will be shown some descriptions of crimes and asked to rate the seriousness of those crimes. This survey is expected to take about 10 minutes to complete.

At the end of the survey, you will receive a unique, randomly generated code. Please copy and paste it on M-Turk in order to complete the HIT and receive your payment.

Participant Requirements

To participate in this study, you must be age 18 or older and a resident of the United States.

Risks

The risks and discomfort associated with participation in this study are no greater than those ordinarily encountered in daily life or during other online activities.

Benefits

There are no direct personal benefits which may reasonably be expected to result from this study.

Compensation & Costs

You will receive \$1.25 as payment for your participation in this study.

There will be no cost to you if you participate in this study.

Confidentiality

By participating in this study, you understand and agree that the data and information gathered during this study may be used by Carnegie Mellon for academic publications. The Federal government offices that oversee the protection of human subjects in research may have access to research records. Carnegie Mellon may be required to disclose your survey responses and other information as required by law, regulation, subpoena or court order. That information may include your IP address. We also collect information about your operating system and web browser.

Right to Ask Questions & Contact Information

If you have any questions about this study, feel free to ask them by contacting the Principal Investigator listed above. If you have questions later, desire additional information, or wish to withdraw your

If you have questions pertaining to your rights as a research participant; or wish to report objections to this study, you should contact the Research Regulatory Compliance Office at Carnegie Mellon University. Email: irb-review@andrew.cmu.edu . Phone: 412-268-1901 or 412-268-5460.

Your participation in this research is voluntary and you may quit the survey at any time.

	Yes	No
I am age 18 or older.	<input type="radio"/>	<input type="radio"/>
I currently reside in the United States.	<input type="radio"/>	<input type="radio"/>
I have read and understand the information above.	<input type="radio"/>	<input type="radio"/>
I want to participate in this research and continue with the survey.	<input type="radio"/>	<input type="radio"/>

Please take the time to read the following important instructions.

On the next three pages, we will show you a list of crimes. On the first page, we will ask you to rate how serious you think each crime is. On the second page, we will ask you to tell us how wrongful you think the crimes are. On the third page, we will ask you to indicate the harmfulness of each crime. At the end of the survey we will ask you a few questions about yourself.


Page 3:

Please use the slider to indicate how serious you think each crime is. If you think a crime is not serious at all, slide the marker to all the way to the left. If you think the crime is extremely serious, move the marker all the way to the right. If you think the crime falls somewhere between the two extremes, slide the marker to a position that best shows how serious you think the crime is.

A person installs monitoring software on another person's computer without permission.

Not at all serious

Extremely serious



A postal employee steals mail containing \$200,000 in checks from dozens of homes.



An employee takes a copy of his employer's confidential customer lists with him without permission when he quits. The customer list is worth \$4,000.



A person breaks into a unguarded railroad car and steals \$200,000 worth of goods.



An IRS employee looks at a celebrity's tax records out of curiosity.



A postal employee steals mail containing \$4,000 in checks from dozens of homes.



A business owner uses his customers' passwords to log into a competitor's web site to view the site design. The design would cost \$200,000 to develop independently.



A bank employee embezzles \$200,000 from his employer



A business owner uses his customers' passwords to log into a competitor's web site to view the site design. The design would cost \$4,000 to develop independently.



A person collects \$200,000 in Social Security payments by using someone else's Social Security Number.



A pharmacist submits false Medicare claims totaling \$4,000.



A person breaks into an unoccupied bank at night and steals \$200,000.



A scammer makes \$200,000 by sending letters that trick dozens of people into sending "processing fees" to receive fake lottery winnings.



A person breaks into an unoccupied post office at night and steals \$200,000.



A person downloads 100,000 email addresses from a company's website with a security flaw. The company spends \$4,000 responding to the incident.



A bank employee embezzles \$4,000 from his employer.













An employee takes a copy of his employer's confidential customer lists with him without permission when he quits. The customer list is worth \$200,000.



A foreign national stows away on a cargo ship headed to the United States.






A person carrying a gun enters a fenced area of a nuclear power plant without permission.	
A person knowingly trespasses in a wildlife refuge.	
A person reads someone else's e-mail without their permission.	
A person collects \$4,000 in Social Security payments by using someone else's Social Security Number.	
A pharmacist submits false Medicare claims totaling \$200,000.	
A scammer makes \$4,000 by sending letters that trick dozens of people into sending "processing fees" to receive fake lottery winnings.	
A passenger tries to bring a weapon through airport security.	
A person breaks into an unoccupied pharmacy and steals \$200,000 worth of drugs.	
A person downloads 100,000 email addresses from a company's website with a security flaw. The company spends \$200,000 responding to the incident.	
A person camps in a national park where camping is not allowed.	



















Note: Order of vignettes is randomized for each participant.

Page 4:

Next, we would like to know how morally wrongful you think each crime is.

Please use the slider to indicate how morally wrong you think it is for a person to commit the crime. If you think it is not morally wrong at all, slide the marker to all the way to the left. If you think committing the crime is extremely morally wrong, slide the marker all the way to the right. If your judgment is somewhere in between, slide the marker to a position that best shows how morally wrong you think it is to commit the crime.

	Not at all wrongful	Extremely wrongful
A person installs monitoring software on another person's computer without permission.		
A postal employee steals mail containing \$200,000 in checks from dozens of homes.		
An employee takes a copy of his employer's confidential customer lists with him without permission when he quits. The customer list is worth \$4,000.		

A person breaks into a unguarded railroad car and steals \$200,000 worth of goods.	
An IRS employee looks at a celebrity's tax records out of curiosity.	
A postal employee steals mail containing \$4,000 in checks from dozens of homes.	
A business owner uses his customers' passwords to log into a competitor's web site to view the site design. The design would cost \$200,000 to develop independently.	
A bank employee embezzles \$200,000 from his employer	
A business owner uses his customers' passwords to log into a competitor's web site to view the site design. The design would cost \$4,000 to develop independently.	
A person collects \$200,000 in Social Security payments by using someone else's Social Security Number.	
A pharmacist submits false Medicare claims totaling \$4,000.	
A person breaks into an unoccupied bank at night and steals \$200,000.	
A scammer makes \$200,000 by sending letters that trick dozens of people into sending "processing fees" to receive fake lottery winnings.	
A person breaks into an unoccupied post office at night and steals \$200,000.	
A person downloads 100,000 email addresses from a company's website with a security flaw. The company spends \$4,000 responding to the incident.	
A bank employee embezzles \$4,000 from his employer.	
An employee takes a copy of his employer's confidential customer lists with him without permission when he quits. The customer list is worth \$200,000.	
A foreign national stows away on a cargo ship headed to the United States.	
A person carrying a gun enters a fenced area of a nuclear power plant without permission.	
A person knowingly trespasses in a wildlife refuge.	
A person reads someone else's e-mail without their permission.	

A person collects \$4,000 in Social Security payments by using someone else's Social Security Number.	
A pharmacist submits false Medicare claims totaling \$200,000.	
A scammer makes \$4,000 by sending letters that trick dozens of people into sending "processing fees" to receive fake lottery winnings.	
A passenger tries to bring a weapon through airport security.	
A person breaks into an unoccupied pharmacy and steals \$200,000 worth of drugs.	
A person downloads 100,000 email addresses from a company's website with a security flaw. The company spends \$200,000 responding to the incident.	
A person camps in a national park where camping is not allowed.	

Note: The order of vignettes is the same as on Page 3.

Page 5:

	Not at all harmful	Extremely harmful
A person installs monitoring software on another person's computer without permission.		
A postal employee steals mail containing \$200,000 in checks from dozens of homes.		
An employee takes a copy of his employer's confidential customer lists with him without permission when he quits. The customer list is worth \$4,000.		
A person breaks into a unguarded railroad car and steals \$200,000 worth of goods.		
An IRS employee looks at a celebrity's tax records out of curiosity.		
A postal employee steals mail containing \$4,000 in checks from dozens of homes.		

A business owner uses his customers' passwords to log into a competitor's web site to view the site design. The design would cost \$200,000 to develop independently.



A bank employee embezzles \$200,000 from his employer



A business owner uses his customers' passwords to log into a competitor's web site to view the site design. The design would cost \$4,000 to develop independently.



A person collects \$200,000 in Social Security payments by using someone else's Social Security Number.



A pharmacist submits false Medicare claims totaling \$4,000.



A person breaks into an unoccupied bank at night and steals \$200,000.



A scammer makes \$200,000 by sending letters that trick dozens of people into sending "processing fees" to receive fake lottery winnings.



A person breaks into an unoccupied post office at night and steals \$200,000.



A person downloads 100,000 email addresses from a company's website with a security flaw. The company spends \$4,000 responding to the incident.



A bank employee embezzles \$4,000 from his employer.



An employee takes a copy of his employer's confidential customer lists with him without permission when he quits. The customer list is worth \$200,000.



A foreign national stows away on a cargo ship headed to the United States.



A person carrying a gun enters a fenced area of a nuclear power plant without permission.



A person knowingly trespasses in a wildlife refuge.



A person reads someone else's e-mail without their permission.



A person collects \$4,000 in Social Security payments by using someone else's Social Security Number.



A pharmacist submits false Medicare claims totaling \$200,000.



A scammer makes \$4,000 by sending letters that trick dozens of people into sending "processing fees" to receive fake lottery winnings.



A passenger tries to bring a weapon through airport security.



A person breaks into an unoccupied pharmacy and steals \$200,000 worth of drugs.



A person downloads 100,000 email addresses from a company's website with a security flaw. The company spends \$200,000 responding to the incident.



A person camps in a national park where camping is not allowed.



Page 6: Thank you

Thank you! We would now like to ask you some questions about yourself.

Cybercrime Attitudes - Misc

Have you personally been the victim of any of the following crimes?

- ☐ Data Breach
- ☐ Credit card fraud
- ☐ Identity theft (other than unauthorized credit card charges) Burglary
- ☐ Fraud (other than credit card fraud)
- ☐ None of the above

If you have been a victim of any of these crimes, can you please provide more details? (Please do not reveal anything in your answer that is private and could identify you.)

Page 7: Demographics

What is your gender?

- ☐ Male
- ☐ Female
- ☐ Non-binary / Transgender / Other _____
- ☐ Prefer not to answer

What is your current age? (Please enter "-1" if you prefer not to answer)

What race or races do you consider yourself to be?

- ☐ White
- ☐ Black or African American
- ☐ American Indian or Alaska Native

- ☐ Asian
- ☐ Native Hawaiian or Pacific Islander
- ☐ Other _____
- ☐ Prefer not to Answer

What political party do you most identify with?

- ☐ Democrat
- ☐ Independent
- ☐ Republican
- ☐ Other
- Prefer not to answer

Where were you born?

[Drop-down list of countries]

Where do you live now?

[Drop-down list of countries]

What is the highest level of education you have completed?

- ☐ Less than High School
- ☐ High School / GED
- ☐ Some College
- ☐ 2-year College Degree
- ☐ 4-year College Degree
- ☐ Masters Degree
- ☐ Professional Degree (JD, MD)
- ☐ Doctoral Degree
- ☐ Prefer not to answer

What is your current work situation?

[Drop-down list of work situations]

Which of the following most closely describes your current occupation?

[Drop-down list of occupations]

Page 8: Open-Ended

Finally, what do you think this study was about?

Page 9: Conclusion

Thank you for taking part in our survey! The purpose of this survey was to measure whether people view certain computer crimes as being more like trespass, burglary, or fraud crimes.

Please click on ">>" to complete your task and receive your unique, randomly generated compensation code. Remember to copy that code and paste it into the Mechanical Turk survey to be paid.

Please let us know if you have any comments about this study.