Journal of Criminal Law and Criminology

Volume 109
Issue 2 *Spring*

Article 4

Spring 2019

Perception Versus Punishment in Cybercrime

James T. Graves

Alessandro Acquisti

Ross Anderson

0091-4169/19/10902-0313
THE JOURNAL OF CRIMINAL LAW & CRIMINOLOGY Vol. 109, No. 2
 Printed in U.S.A.

CRIMINOLOGY

PERCEPTION VERSUS PUNISHMENT IN CYBERCRIME

JAMES T. GRAVES, ALESSANDRO ACQUISTI & ROSS ANDERSON*

TABLE OF CONTENTS

* James T. Graves is a Ph.D. Candidate at Carnegie Mellon University and Staff Attorney in the Institute for Public Representation at Georgetown University Law Center. Alessandro Acquisti is Professor of Information Technology and Public Policy at Carnegie Mellon University's Heinz College. Ross Anderson is Professor of Security Engineering at the University of Cambridge. This work was partially funded by the Department of Homeland Security Science and Technology Directorate, Cyber Security Division, Broad Agency Announcement 11.02; the Government of Australia; and SPAWAR Systems Center Pacific, via contract number N66001-13-C-0131. Portions of this work were also supported by NSF IGERT grant DGE-0903659. In addition, Acquisti gratefully acknowledges support from the Carnegie Corporation of New York via an Andrew Carnegie Fellowship. For a list of Acquisti's additional grants and funding sources, please visit www.heinz.cmu.edu/~acquisti/cv.htm. This work represents the position of the authors and not that of the aforementioned agencies.

I. INTRODUCTION

The U.S. Computer Fraud and Abuse Act (CFAA)[1] is not a popular law.[2] 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.[3]

---

[1] 18 U.S.C. § 1030 (2018).

[2] *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 [http://perma.cc/SW7Y-YU2Q] (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 [http://perma.cc/WTB4-2UQT] (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 [http://perma.cc/YZ33-GQ6A] (arguing that the CFAA criminalizes online activism).

[3] *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*

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.[4]

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 rapidly as (easily inflated) losses increase.[5] 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.[6] Criminal law codifies social norms, which manifest as perceptions that can be empirically measured.[7] More generally, public opinion influences policymaking.[8] Criminal codes "reflect through the state

---

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).

[4] *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 [http://perma.cc/2HGS-S49P] (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 [http://perma.cc/LU4P-2MG5] (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 [http://perma.cc/J2SF-2BWT] (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").

[5] *See*, *e.g.,* Granick, *supra* note 3, at 211.

[6] *See*, *e.g.*, Michael O'Connell & Anthony Whelan, *Taking Wrongs Seriously*, 36 BRIT. J. CRIMINOLOGY 299, 299 (1996); Part B, *infra*.

[7] *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).

[8] *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.

legislature's deliberations and actions some understanding, however dim and remote, of what 'the public' deems appropriate for the crimes in question."[9] Although public perceptions of the criminal justice system are flawed,[10] 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.

We report 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. We 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.[11] 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.

Our 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 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 piece proceeds as follows. Part 0 provides background information. In Part II.A, we 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 Part II.B, we summarize previous work on crime seriousness. Part III presents the methodology,

---

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).

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

[10] *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).

[11] *See infra* note 97 and accompanying text.

model, and results of our between-subjects experiments. Part IV presents our factorial vignette survey experiment. Part 0 discusses the implications of our results and concludes.

II. BACKGROUND

A. 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.[12] The sentencing guidelines determine the recommended sentencing range based on aspects of both the crime and relevant conduct.[13] The rest of this section discusses how various factors of a CFAA crime affect maximum and recommended sentences.

1. Maximum Sentences

The CFAA criminalizes six types of conduct as "computer crime."[14] In general terms, these are (1) obtaining information,[15] (2) accessing government computers,[16] (3) committing computer fraud,[17] (4) causing damage with or to a computer,[18] (5) trafficking in passwords,[19] and (6) extorting money by threatening to obtain information or damage a computer.[20] Table 1 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 extortion, which have maximum sentences of five years for a first offense,[21] and accessing

---

[12] *See* 18 U.S.C. § 1030(c).

[13] *See* U.S. SENTENCING GUIDELINES MANUAL §§ 2B1.1, 2B2.3, 2M3.2, 2X1.1 (U.S. Sentencing Comm'n 2015).

[14] For in-depth discussions of the CFAA, *see generally* Computer Crime & Intellectual Prop. Section Crim. Div., DEP'T OF JUST., PROSECUTING COMPUTER CRIMES, https://www.justice.gov/sites/default/files/criminal-ccips/legacy/2015/01/14/ccmanual.pdf [http://perma.cc/8N8N-2FU9]; Kerr, *Cybercrime's Scope*, *supra* note 2.

[15] 18 U.S.C. § 1030(a)(1)–(2).

[16] *Id.* § 1030(a)(3).

[17] *Id.* § 1030(a)(4).

[18] *Id.* § 1030(a)(5).

[19] *Id.* § 1030(a)(6).

[20] *Id.* § 1030(a)(7).

[21] *Id.* § 1030(c).

Table 1: 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.

national security information, with a maximum sentence of ten years for a first offense.[22]

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."[23]

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.[24] If the offender intentionally caused any of the forms of damage listed above, the maximum sentence increases to ten years.[25] And if the offender intentionally caused serious bodily injury or death, the maximum sentence increases to twenty years or life, respectively.[26]

If the data obtained in a cybercrime includes "a means of identification of another person," the crime can be charged under the identity theft

[22] *Id.* § 1030(a).
[23] *Id.* § 1030(a)(2), (c)(2)(B).
[24] *Id.* § 1030(c)(4)(A)(i).
[25] *Id.* § 1030(c)(4)(B)(ii).
[26] *Id.* § 1030(c)(4)(E)–(F).

statutes.[27] A conviction for identity theft carries a maximum sentence of five years.[28] 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.[29]

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.[30] *Motive* is reflected in an increased maximum sentence of five years for obtaining information for purposes of commercial advantage or financial gain.[31] The *consequences* of a CFAA crime can increase sentences through the $5000 loss threshold in certain subsections[32] and through maximum sentences that grow longer as damage increases to include physical injury, serious bodily injury, or death.[33] By *context*, we 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.[34] 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.[35] If an offender accessed classified national security information, the maximum sentence for a first offense increases to ten years.[36]

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,[37] the guidelines are intended to "provide certainty and fairness in meeting the purposes of sentencing, avoiding

---

[27] *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."

[28] *Id.* § 1028(b)(2)(B).

[29] *Id.* § 1028A(a)(1).

[30] *Id.* § 1030(c)(4)(A)(i), (B)(i).

[31] *Id.* § 1030(c)(2)(B).

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

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

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

[35] *Id.* §§ 1028(b)(2)(B), 1028A(a)(1).

[36] *Id.* § 1030(a)(1), (c)(1).

[37] 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).

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[.]"[38]

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 conviction, a court determines the offense level and criminal history category then consults the table reproduced in this article in Table 7. 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."[39] 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.[40] 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.[41] The base offense level for most CFAA crimes is six.[42] Computer extortion has a base offense level of eighteen, and unauthorized access to national security information carries a base offense level of thirty.[43]

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.[44] That increase is roughly equivalent to an additional 8 to 10 years in prison

---

[38] 28 U.S.C. § 991(b)(1)(B).

[39] U.S. SENTENCING GUIDELINES MANUAL § 4B1.1 (U.S. Sentencing Comm'n 2016).

[40] *Id.* app. A (U.S. Sentencing Comm'n 2016) (indexing statutes to sentencing guidelines sections).

[41] *Id.*

[42] *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.

[43] *Id.* §§ 2B3.2, 2M3.2.

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

(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.[45] For example, Aaron Swartz allegedly downloaded 4.8 million articles that cost $19 each to download from JSTOR.[46] 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 sentence length—for a crime with ten or more victims or at least one victim who suffered "substantial financial hardship."[47] 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.[48]

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"[49] and a

---

[45] *See, e.g.*, Granick, *supra* note 3, 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.").

[46] 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/ [http://perma.cc/4YMX-V7SV] (last visited Dec. 14, 2016).

[47] 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).

[48] *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.

[49] *Id.* § 2B1.1(b)(11).

separate two-point increase if the offense involved "an intent to obtain personal information" or "unauthorized public dissemination of personal information."[50] The penalty for accessing national defense information increases the base offense level from thirty to thirty-five if the information was classified Top Secret.[51]

Enhancements may also be based on the target of a crime (what we 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.[52] An additional six-point enhancement applies if the offense caused "substantial disruption of a critical infrastructure."[53]

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,[54] acceptance of responsibility,[55] use of a "special skill"[56] or "sophisticated means,"[57] and motivation.[58] 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.[59]

B. CRIMINOLOGICAL STUDIES OF CRIME SERIOUSNESS

Criminologists have been studying perceptions of crime seriousness for nearly a hundred years.[60] In 1922, Willis Clark asked 100 people to "grade"

---

[50] *Id.* § 2B1.1(b)(17).

[51] *Id.* § 2M3.2.

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

[53] *Id.* § 2B1.1(b)(18)(iii).

[54] *See id.* §§ 3B1.1–3B1.5.

[55] *See id.* § 3E1.1.

[56] *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.

[57] *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.

[58] *See id.* §§ 3A1.1, 3A1.4.

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

[60] 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:*

on a scale from one to ten the seriousness of 148 acts of delinquency committed by schoolboys.[61] Categorizing these acts into different types (truancy, stealing, "incorrigibility," "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,[62] Sellin and Wolfgang are generally credited with the pioneering empirical research.[63] They sought to create a data-based index of delinquency that could be used to evaluate the effectiveness of efforts to combat juvenile crime.[64] 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.[65] They therefore conducted the first rigorous and comprehensive empirical study of attitudes towards crime by surveying judges, police, and college students in Philadelphia to come up with rankings for 141 different offenses.[66] Other scholars soon replicated and extended their work.[67]

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 several 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?

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).

[61] 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).

[62] *See* Sweeten, *supra* note 60, at 535–37.

[63] *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 60, at 37 ("The study of perceptions of crime seriousness was introduced by Sellin and Wolfgang.").

[64] THORSTEN SELLIN & MARVIN E. WOLFGANG, THE MEASUREMENT OF DELINQUENCY 1 (1964).

[65] *Id.* at 6.

[66] *Id.* at 241–58.

[67] *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 63, at 224 (surveying households in Baltimore to obtain ratings of a set of 140 crimes).

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.[68] 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.[69] Other respondents appeared to ignore moral wrongfulness entirely, judging crimes solely on the harm done.[70]

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.[71] Stephen Blum-West looked at eight dimensions: bodily harm, economic damage, emotional damage, potential for harm, intent, purpose, motive, and fair play.[72]

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

---

[68] 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).

[69] Warr, *supra* note 68, at 802–08.

[70] *Id.* at 810–15.

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

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

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 scenario[73]—and indeed we use a similar approach in one of our 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.[74] 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.[75] 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.[76]

Rossi, Simpson, and Miller were among the first to apply the factorial vignette methodology to perceptions of crime seriousness.[77] 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 program to print booklets of 50 vignettes each that respondents rated on paper. The

---

[73] *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).

[74] *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.

[75] *See* Rossi et al., *supra* note 9, at 68–69.

[76] *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.

[77] *Id.* at 62.

rating task was to mark an unnumbered line answering "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.[78] 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,[79] and Guillermina Jasso discusses it in depth in the context of measuring judgments using factorial vignette surveys.[80]

Other work in studying crime seriousness has focused on particular types of crime. For example, criminologists have studied perceptions of white–collar crimes,[81] environmental crimes,[82] and "small" crimes.[83] 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.[84] Although white–collar crimes may be similar to computer crimes, to our knowledge no one has analyzed how the features of cybercrimes affect perceptions.

---

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

[79] Rossi et al., *supra* note 9, at 81–89.

[80] *See* Jasso, *supra* note 74, at 388–403.

[81] *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).

[82] *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).

[83] *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).

[84] *See* Cullen et al., *supra* note 81, at 83, 92–94; Dodge et al., *supra* note 73, at 412 (2013).

III. STUDY I: BETWEEN-SUBJECTS EXPERIMENTS

To understand how features of cybercrimes affect individuals' perceptions, we conducted two human-subjects studies whose methodologies complement each other (see Introduction). 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 Part IV.

A. METHODOLOGY

1. Research Questions

We designed six between-subjects experiments, randomly assigning each subject to one experimental condition. In each experiment, we 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. We chose this vignette for a number of reasons. The data-breach scenario is a common one that we believe is readily understandable by most people.[85] It also lends itself to manipulation of the attributes of interest (scope, context, motivation, etc.) while holding other attributes reasonably constant.

We focus on six aspects of cybercrime likely to influence perceptions of wrongfulness or harmfulness. Five of them are, as discussed in Part A, directly relevant to sentencing: (1) scope, (2) motivation, (3) consequences, (4) context, and (5) the type of data affected. We 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, we 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.*

---

[85] *See* Data Breaches, PRIVACY RIGHTS CLEARINGHOUSE, http://www.privacyrights.org/data-breach [http://perma.cc/U4LW-PEE8] (last visited Aug. 31, 2018) (stating that over 11 billion data records have been affected in over 8,000 data breaches since 2005).

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.*

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

2. Design

Study I consisted of six between-subjects online survey experiments. Within each experiment, we randomly assigned participants 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. We were 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. We tackled these issues by specifying consequences whenever possible and by stating in the vignette that the perpetrator of the data breach in our scenario did not release the data he downloaded. This had the desirable side effect of limiting extreme "ceiling" effects in the responses to our 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, we 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. We selected the first three questions in accordance with previous research on the factors of crime seriousness.[86] 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?"[87]
>
> "How clever was Mr. Tom Smith?"

---

[86] *See, e.g.*, Warr, *supra* note 68, at 796.

[87] In the Context experiment, the "Acme Insurance Company" was replaced by "ACR."

"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. We 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. We used the fifteen-question Concern for Information Privacy (CFIP) scale.[88] We 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.

We ran ordered probit regressions on each variable of interest. Regressions included controls for demographics, memory check correctness, and privacy attitudes. We treated the demographic variables for gender, country of birth, age, education, occupation, work situation, and the memory check variables as categorical variables. We 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.

---

[88] *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).

B. THEORETICAL MODEL

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

$$Y = \beta_0 + \beta_1 X + \sum \gamma_q Z_q + \varepsilon$$

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).

Our 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.

C. RESULTS

For each experiment, we 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 article. The demographics and data quality of MTurk experiments have been extensively studied in multiple experimental contexts.[89] Several studies have shown that recruitment for online studies through MTurk can lead to more representative samples and better data quality than studies using other "convenience"

[89] *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 [http://perma.cc/5ZJ2-K4T5] (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 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 [http://perma.cc/C5EY-DJXK] (using MTurk to replicate a telephone survey).

samples such as university students.[90] Peer and his co-authors found that reputation alone is often enough to ensure sufficient data quality in MTurk studies.[91] Another study showed that MTurkers paid more attention to instructions than did traditional subject pool samples.[92]

Our MTurk job description asked people to take "a short survey on crime." We recruited a total of 2,635 participants in October through December 2013. We screened potential participants to exclude anyone who had participated in our crime seriousness experiments from participating in subsequent experiments in this series. We 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 $p < 0.05$. We account for these variables (and all other demographic variables) in our regressions.

---

[90] *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).

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

[92] 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 2: 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	0.194	0.076	-0.028	0.970***	0.015	0.008		239
		(0.142)	(0.145)	(0.148)	(0.145)	(0.151)	(0.153)	(0.143)		
Scope	log(Records)	0.070**	0.078**	0.159***	0.107***	0.135***	0.064*	0.057*		583
		(0.27)	(0.026)	(0.028)	(0.026)	(0.031)	(0.026)	(0.025)		
Motivation	Student (v. Profiteer)	-0.878***	-0.327*	-0.596***	-0.793***	0.201	0.034	0.217	-0.051	361
		(0.151)	(0.148)	(0.150)	(0.145)	(0.141)	(0.141)	(0.141)	(0.147)	
Motivation	Activist (v. Profiteer)	-0.795***	-0.279	-0.538***	-0.497***	0.130	0.100	0.191	-0.294	361
		(0.150)	(0.145)	(0.152)	(0.147)	(0.154)	(0.145)	(0.152)	(0.159)	
Conseq.	Acme (v. Low)	0.179	0.407***	0.083	0.338**	0.147	-0.009	-0.123	-0.020	479
		(0.123)	(0.122)	(0.119)	(0.123)	(0.137)	(0.140)	(0.116)	(0.118)	
Conseq.	Customers (v. Low)	0.042	0.377**	0.131	0.236*	0.093	0.040	0.112	-0.125	479
		(0.125)	(0.120)	(0.121)	(0.118)	(0.138)	(0.151)	(0.126)	(0.124)	
Co-Resp.	Patched (v. Not)	0.133	0.102	0.157	0.074	0.087	-0.370*	0.423***	-0.184	276
		(0.136)	(0.136)	(0.133)	(0.132)	(0.151)	(0.164)	(0.128)	(0.136)	
Context	Gov't (v. Bank)	-0.055	0.013	-0.027	-0.030	0.147	-0.121	0.152	-0.023	502
		(0.119)	(0.121)	(0.125)	(0.116)	(0.139)	(0.142)	(0.118)	(0.116)	
Context	Non-Profit (v. Bank)	0.048	-0.029	-0.222	0.030	0.099	-0.208	-0.361**	-0.185	502
		(0.123)	(0.124)	(0.122)	(0.121)	(0.140)	(0.155)	(0.120)	(0.121)	

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Table 2 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 our 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, were not statistically significantly different. 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 our other experiments. Although we 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. We 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 we would expect, but that the two conditions are not as similar as we 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 had 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 our participants to cybercrimes committed by activists versus cybercrimes committed for profit. The former were considered significantly less

Table 3: 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.

blameworthy, and deserving significantly lighter sentences—contrary to the position sometimes taken by U.S. prosecutors.

Table 3 shows pairwise correlations between each dependent variable across all six between-subjects experiments in Study I.[93] 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.[94] 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, our 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 our 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.

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.

[93] Correlation matrices for each study do not differ meaningfully from the aggregate.

[94] *See* Warr, *supra* note 68, at 818–20.

IV. STUDY II: FACTORIAL VIGNETTE SURVEY EXPERIMENT

We followed the between-subjects experiments with an experiment using factorial vignette survey methodology. As discussed in Part B, factorial vignette surveys are commonly used to study beliefs and normative judgments.[95] 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."

We decided to supplement our between-subjects experiments with a factorial vignette survey experiment for several 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 our 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.[96]

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 our case) to avoid fatigue, the statistical power of this analysis is limited.

Third, the different methodology lets us test the robustness of our 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 we asked participants to do.

Finally, the factorial vignette survey is a known methodology that has been used already in the literature on crime seriousness.[97]

---

[95] *See* KATRIN AUSPURG & THOMAS HINZ, FACTORIAL SURVEY EXPERIMENTS 13–15 (2015); Jasso, *supra* note 74, at 338–39; Rossi et al., *supra* note 9, at 62.

[96] *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).

[97] *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 74 (using

A. METHODOLOGY

1. Research Questions

Our research questions are driven by the goals listed in Section 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. We conjectured that the same would be true when all were manipulated in the same study.

2. Design

The design for this study consisted of a factorial vignette survey experiment. We presented each participant with twenty-five vignettes describing a cybercrime scenario.[98] Each was structured as a paragraph describing the facts followed by a list of the factors that varied from one vignette to another.[99] 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 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.

---

the factorial survey methodology to study perceptions of five types of crimes); Rossi et al., *supra* note 9 (using the factorial survey methodology to study perceptions of fifty crimes).

[98] We 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. We therefore scaled back to 25 vignettes.

[99] Adopting a variation of Jasso's terminology, we 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 74, at 340–41 (2006).

> 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*.

We 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 we 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. We did not prevent other combinations that some participants might have found implausible, such as an organization suffering $10 million in losses from the breach of 10 e-mail addresses (a combination that occurred 29 times in our data set). Treating the numerical factors *$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 we presented a slider with the rating task asking participants to evaluate the sentence imposed. We limited the rating

Figure 1: Factorial instrument rating task slider

task to one question because of research showing that undesirable method effects increase when participants are asked multiple questions after each vignette.[100] The slider was anchored at each end with "Much too low" at the left and "Much too high" at the right. We set the marker on the slider to a starting position in the middle of the scale. We left the slider 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.[101] Other research shows that adding numeric labels to a slider leads to increased rounding of responses.[102] Figure 1 shows the slider scale that we used.

We used a slider bar to approximate the real-number scale used in some previous factorial vignette surveys.[103] The slider widget we used 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." We normalized this to a 0 to 100 scale with 100 corresponding to a response that the punishment should have been higher—i.e., we reversed the scale as presented. We 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

---

[100] *See* 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), https://www.iser.essex.ac.uk/research/publications/working-papers/iser/2012-01 [http://perma.cc/25D9-CYZS].

[101] *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/ [http://perma.cc/U2QT-778U].

[102] *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).

[103] *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 9, at 66–67 (1985).

they thought the study was about and an optional question in which participants could enter comments about the study.

We 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, we treated gender, country of birth, age category, education, occupation, work situation, and the memory check variables as categorical variables. We treated as continuous variables the extent to which participants had been affected by cybercrime or privacy invasions and the extents to which they use fake personal information and are aware of media coverage of data misuse.

B. THEORETICAL MODEL

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

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

where $i = 1 \ldots n$ indexes the vignettes, $j = 1 \ldots m$ 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.[104] This assumption simplifies the model and lets us understand beliefs in the aggregate.

C. RESULTS

We 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 article. We 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

[104] *See* Jasso, *supra* note 74, at 350–51.

that her age was under 18, a total of 223 responses remained (47% women; median age category 25–34).[105]

Table 4 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 our attention check question correctly.[106]

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. The next highest effect is the type of data. This is somewhat surprising because the type of data was not a manipulation that produced statistically significant differences 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

---

[105] 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).

[106] In a regression excluding 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.

Table 4: 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)
N	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.

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 our 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.

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

V. DISCUSSION

A. 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 4 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 4.[107] Perhaps respondents used the length of the data type as a heuristic. Unfortunately, because the length of our data type

[107] Reading down the column: 7.8, 11.6, 12.4, 5.9, and 5.3.

descriptions and the sensitivity of the data listed are not independent, it is impossible to disentangle their effects in our results.

B. 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. Our main factorial regression equation predicts that increasing the organization's loss from $1,000 to $10,000,000 corresponds to a 5.9-point increase in severity rating (on a 100-point scale). The same change in dollar amount would lead to a 20-point increase in offense level in the 2016 Sentencing Guidelines,[108] enough to bump the presumptive sentencing range for a first offense with no other enhancements from 0–6 months to 63–78 months.[109] For comparison, the coefficient on *$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 our experiment rate as appropriate.

Motivation was much more important in our 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 our 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.

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,[110] or government computers.[111]

---

[108] *See* U.S. SENTENCING GUIDELINES MANUAL § 2B1.1(b)(1).

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

[110] 18 U.S.C. § 1030(a)(2)(A)–(B).

[111] *Id.* § 1030(a)(3).

Table 5: 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.

Table 5 lists the effect of offense factors on perceptions and sentencing. For example, our 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).

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 we used in our 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. Our 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 our 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 our 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 gain.[112] Also, the value of the records Tom obtained may well be worth more than $5000.[113] The base offense level under section 2B1.1 would be 6.[114] The enhancements for using special skill[115] or sophisticated means,[116] 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.[117] 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 our 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,[118] 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."[119] 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

---

[112] *See id.* § 1030(c)(2)(B)(i).

[113] 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/ [http://perma.cc/35H7-L57Q]. 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 [http://perma.cc/JCH4-J4Z9]; 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/ [http://perma.cc/4WAM-X4RJ]. Even at a nickel per record, however, a set of 100,000 records would be worth $5000.

[114] U.S. SENTENCING GUIDELINES MANUAL § 2B1.1(a).

[115] *Id.* § 3B1.3.

[116] *Id.* § 2B1.1(b)(10)(C).

[117] *Id*. § 2B1.1(b)(17).

[118] 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 [http://perma.cc/4V5J-RPTQ].

[119] 18 U.S.C. § 1030(c)(2)(B)(ii).

information" as defined in the guidelines),[120] 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 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.[121] 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.

Two lessons can be gleaned from these examples (which Table 6 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

---

[120] 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 (U.S. Sentencing Comm'n 2016).

[121] 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).

Table 6: Sentencing examples for the factorial scenario

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

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.

Apart from the language about gaining access to the company's internal network, our 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.[122] Auernheimer was convicted of conspiracy and identity fraud.[123] He received a sentence of 41 months that was overturned on jurisdictional grounds.[124]

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.[125] 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

[122] *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/ [http://perma.cc/TU6G-YD44].

[123] Judgment in a Criminal Case, United States v. Auernheimer, No. 2:11-cr-470 (D.N.J. Mar. 19, 2013), ECF No. 92.

[124] *Id.*

[125] Letter from United States, *Auernheimer*, No. 2:11-cr-470 (D.N.J. Mar. 15, 2013), ECF No. 89.

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.[126] 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.[127]

As mentioned in section 2, most CFAA offenses are sentenced under section 2B1.1 of the guidelines, which covers economic crimes such as fraud and larceny. Our results support arguments that this is a poor fit.[128] 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 our respondents do rate as important, such as motive, type of data, and scope, are barely factors in 2B1.1.

C. LIMITATIONS AND OPPORTUNITIES FOR FURTHER RESEARCH

We emphasize, as Rossi, Simpson, and Miller did in 1985,[129] that we do not claim that sentences should be determined by public opinion. As we mentioned at the beginning of this article, lay opinions of sentencing are subject to biases, lack of information, and misperceptions. However, these perceptions do inform public policy decisions. When perceptions are wildly out of line with sentencing mechanisms, it is worth asking whether those mechanisms truly achieve public policy objectives. Furthermore, our measurement of perceptions is focused on the relative importance of various factors rather than on the comparison of total sentences.

The experiments we 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 our work would be to compare different types of cybercrime. In addition, we intend to study how cybercrimes are perceived in comparison with similar real-world crimes.

Another limitation of this work is that it ignores many victim and offender characteristics, other than the offender's cleverness. The victims in

---

[126] 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 45, at 1557–58.

[127] *See* 18 U.S.C. § 1030(c)(2)(b)(iii).

[128] *See* Kerr, *supra* note 45, at 1554–56.

[129] Rossi et al., *supra* note 9, at 61.

our 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 our results, assuming these unobserved characteristics and participant assumptions about them were distributed randomly, it is possible that the effects we do measure are smaller than those we chose to ignore.

Because we use MTurk for our 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 our 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 we describe in this article support the argument that most computer crimes should not be sentenced as fraud crimes, our results say nothing about whether trespass is the correct analogue. Computer crimes also have features of burglary, for example. Future work might explore this further.

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

We 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,"[130] it is reasonable to ask whether those reflections are distorted. Our research suggests that they are.

---

[130] *Id.* at 59–60.

APPENDIX A: U.S. SENTENCING GUIDELINES TABLE

Table 7: U.S. Sentencing Guidelines Sentencing Table

		Criminal History Category (Criminal History Points)					
	Offense Level	**I (0 or 1)**	**II (2 or 3)**	**III (4, 5, 6)**	**IV (7, 8, 9)**	**V (10, 11, 12)**	**VI (13 or more)**
	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
Zone A	**5**	0-6	0-6	1-7	4-10	6-12	9-15
Zone A	**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
	9	4-10	6-12	8-14	12-18	18-24	21-27
Zone B	**10**	6-12	8-14	10-16	15-21	21-27	24-30
Zone B	**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
Zone C	**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
	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
Zone D	**28**	78-97	87-108	97-121	110-137	130-162	140-175
Zone D	**29**	87-108	97-121	108-135	121-151	140-175	151-188
Zone D	**30**	97-121	108-135	121-151	135-168	151-188	168-210
Zone D	**31**	108-135	121-151	135-168	151-188	168-210	188-235
Zone D	**32**	121-151	135-168	151-188	168-210	188-235	210-262
Zone D	**33**	135-168	151-188	168-210	188-235	210-262	235-293
Zone D	**34**	151-188	168-210	188-235	210-262	235-293	262-327
Zone D	**35**	168-210	188-235	210-262	235-293	262-327	292-365
Zone D	**36**	188-235	210-262	235-293	262-327	292-365	324-405
Zone D	**37**	210-262	235-293	262-327	292-365	324-405	360-life
Zone D	**38**	235-293	262-327	292-365	324-405	360-life	360-life
Zone D	**39**	262-327	292-365	324-405	360-life	360-life	360-life
Zone D	**40**	292-365	324-405	360-life	360-life	360-life	360-life
Zone D	**41**	324-405	360-life	360-life	360-life	360-life	360-life
Zone D	**42**	360-life	360-life	360-life	360-life	360-life	360-life
Zone D	**43**	life	life	life	life	life	life

APPENDIX B: REGRESSION TABLES FOR BETWEEN-SUBJECTS EXPERIMENTS (STUDY I)

Table 8: Ordered probit marginal effects for the Type of Data experiment

	Wrongful	Harmful	Serious	Harsh	Sensitive	Respons.	Clever
Medical data	-0.104	0.194	0.076	-0.028	0.970***	0.015	0.008
	(0.142)	(0.145)	(0.148)	(0.145)	(0.151)	(0.153)	(0.143)
Female	0.435**	0.259	0.348*	0.084	0.349*	0.416**	-0.027
	(0.147)	(0.156)	(0.153)	(0.158)	(0.162)	(0.150)	(0.143)
US birth	-0.209	0.141	0.322	0.040	0.521	-0.354	0.177
	(0.227)	(0.348)	(0.295)	(0.293)	(0.326)	(0.447)	(0.214)
CFIP score	0.563***	0.197	0.304**	0.295**	0.501***	0.281*	0.235
	(0.110)	(0.116)	(0.104)	(0.102)	(0.117)	(0.116)	(0.131)
Freq. aff by cybercrime	-0.016	-0.003	-0.127	-0.081	-0.142	0.122	-0.301*
	(0.132)	(0.110)	(0.117)	(0.130)	(0.131)	(0.095)	(0.131)
Fake personal info	-0.010	-0.032	-0.046	-0.093	-0.083	0.040	-0.018
	(0.061)	(0.060)	(0.056)	(0.056)	(0.063)	(0.061)	(0.058)
Media awareness	-0.083	-0.026	0.009	0.016	0.033	0.077	0.064
	(0.051)	(0.049)	(0.049)	(0.044)	(0.051)	(0.051)	(0.053)
AC: Data	0.490	0.396	0.117	-0.106	0.069	-0.660*	-0.204
	(0.257)	(0.297)	(0.252)	(0.248)	(0.282)	(0.307)	(0.239)
AC: Context	-0.287	-0.296	-0.474**	-0.306*	-0.343*	0.100	0.154
	(0.166)	(0.179)	(0.164)	(0.150)	(0.163)	(0.173)	(0.165)
AC: Scope	-0.379	-0.564**	-0.502*	-0.165	-0.410	0.256	-0.127
	(0.194)	(0.209)	(0.212)	(0.184)	(0.229)	(0.196)	(0.208)
N	239	239	239	239	239	239	239
pseudo R^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 9: Ordered probit regression results for the Scope experiment

	Wrongful	Harmful	Serious	Harsh	Sensitive	Respons.	Clever
log(Num. Records)	0.070**	0.078**	0.159***	0.107***	0.135***	0.064*	0.057*
	(0.027)	(0.026)	(0.028)	(0.026)	(0.031)	(0.026)	(0.025)
Female	0.186	0.045	-0.014	0.109	0.240*	-0.145	0.096
	(0.097)	(0.095)	(0.095)	(0.092)	(0.110)	(0.094)	(0.093)
US birth	-0.249	0.028	-0.296	-0.309	-0.210	-0.033	-0.435
	(0.194)	(0.211)	(0.159)	(0.207)	(0.272)	(0.234)	(0.234)
CFIP score	0.361***	0.242***	0.381***	0.241***	0.628***	0.210**	0.261***
	(0.067)	(0.070)	(0.071)	(0.067)	(0.081)	(0.069)	(0.065)
Freq. aff by cybercrime	-0.095	-0.072	-0.187**	-0.102	-0.185*	-0.023	-0.025
	(0.063)	(0.060)	(0.063)	(0.062)	(0.076)	(0.061)	(0.063)
Fake personal info	0.049	-0.045	-0.019	-0.013	-0.032	0.017	0.063
	(0.040)	(0.038)	(0.038)	(0.037)	(0.045)	(0.039)	(0.039)
Media awareness	-0.044	-0.028	-0.032	-0.027	-0.036	0.047	-0.006
	(0.032)	(0.032)	(0.031)	(0.029)	(0.036)	(0.033)	(0.031)
AC: Data	-0.020	-0.133	-0.173	-0.121	0.038	0.010	0.334*
	(0.138)	(0.137)	(0.134)	(0.134)	(0.162)	(0.139)	(0.142)
AC: Context	0.028	-0.010	-0.082	-0.159	0.276	0.107	0.163
	(0.136)	(0.144)	(0.119)	(0.132)	(0.162)	(0.131)	(0.141)
AC: Scope	0.104	-0.031	0.072	0.262*	0.030	-0.056	0.216
	(0.126)	(0.133)	(0.127)	(0.130)	(0.151)	(0.128)	(0.140)
N	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 10: Ordered probit regression results for the Motivation experiment (vs. Profiteer)

	Wrongful	Harmful	Serious	Harsh	Pot. Harm	Sensitive	Respons.	Clever
Student	-0.878***	-0.327*	-0.596***	-0.793***	-0.051	0.201	0.034	0.217
	(0.151)	(0.148)	(0.150)	(0.145)	(0.150)	(0.141)	(0.141)	(0.147)
Activist	-0.795***	-0.279	-0.538***	-0.497***	-0.294	0.130	0.100	0.191
	(0.150)	(0.145)	(0.152)	(0.147)	(0.159)	(0.154)	(0.145)	(0.152)
Female	0.035	-0.037	0.056	-0.051	0.068	-0.106	0.364**	0.001
	(0.121)	(0.123)	(0.126)	(0.128)	(0.129)	(0.121)	(0.124)	(0.119)
US birth	-0.088	0.078	-0.050	0.335	0.042	-0.268	0.053	-0.339
	(0.212)	(0.259)	(0.225)	(0.252)	(0.274)	(0.234)	(0.318)	(0.247)
CFIP score	0.238**	0.181	0.295**	0.223*	0.255**	0.341***	0.140	0.371***
	(0.090)	(0.094)	(0.092)	(0.097)	(0.088)	(0.087)	(0.087)	(0.085)
Freq. aff by cybercrime	0.084	-0.047	0.114	-0.014	0.050	0.011	-0.121	-0.044
	(0.093)	(0.085)	(0.092)	(0.098)	(0.091)	(0.095)	(0.090)	(0.095)
Fake personal info	0.003	-0.007	0.052	-0.007	0.027	-0.029	0.059	-0.045
	(0.053)	(0.051)	(0.053)	(0.052)	(0.052)	(0.052)	(0.051)	(0.053)
Media awareness	0.009	0.100*	0.053	0.033	0.100*	0.026	0.030	-0.029
	(0.044)	(0.047)	(0.045)	(0.043)	(0.047)	(0.042)	(0.044)	(0.042)
AC: Data	-0.313**	-0.115	-0.220	-0.285*	-0.223	-0.510***	0.081	-0.002
	(0.121)	(0.131)	(0.121)	(0.128)	(0.138)	(0.126)	(0.130)	(0.135)
AC: Context	0.058	0.205	0.031	0.093	-0.170	0.032	0.250	0.192
	(0.155)	(0.151)	(0.159)	(0.155)	(0.157)	(0.156)	(0.160)	(0.159)
AC: Scope	0.039	-0.113	0.091	-0.042	-0.079	0.110	-0.079	0.205
	(0.138)	(0.129)	(0.133)	(0.139)	(0.142)	(0.135)	(0.130)	(0.142)
AC: Motivation	-0.208	-0.140	-0.244	-0.234	-0.327	-0.567**	-0.126	-0.014
	(0.179)	(0.192)	(0.177)	(0.170)	(0.188)	(0.178)	(0.174)	(0.189)
N	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 11: Ordered probit regression results for the Consequences experiment (vs. Low condition)

	Wrongful	Harmful	Serious	Harsh	Pot. Harm	Sensitive	Respons.	Clever
Acme High	0.179	0.407***	0.083	0.338**	0.147	-0.009	-0.123	-0.020
	(0.123)	(0.122)	(0.119)	(0.123)	(0.137)	(0.140)	(0.116)	(0.118)
Customers High	0.042	0.377**	0.131	0.236*	0.093	0.040	0.112	-0.125
	(0.125)	(0.120)	(0.121)	(0.118)	(0.138)	(0.151)	(0.126)	(0.124)
Female	0.157	0.113	0.163	0.150	0.261*	0.201	0.129	0.089
	(0.106)	(0.103)	(0.101)	(0.101)	(0.116)	(0.122)	(0.106)	(0.103)
US birth	0.067	-0.116	0.157	0.116	0.008	-0.130	0.071	-0.096
	(0.241)	(0.216)	(0.241)	(0.240)	(0.218)	(0.269)	(0.287)	(0.213)
CFIP score	0.212**	0.168*	0.294***	0.167*	0.417***	0.650***	0.222**	0.119
	(0.076)	(0.082)	(0.078)	(0.080)	(0.101)	(0.104)	(0.074)	(0.078)
Freq. aff by cybercrime	-0.021	-0.002	-0.034	0.007	-0.015	-0.099	0.010	0.020
	(0.079)	(0.076)	(0.078)	(0.077)	(0.088)	(0.097)	(0.075)	(0.073)
Fake personal info	-0.108*	-0.054	-0.094*	-0.115**	-0.026	0.000	0.091*	0.017
	(0.043)	(0.043)	(0.043)	(0.041)	(0.052)	(0.048)	(0.043)	(0.041)
Media awareness	-0.047	0.028	-0.029	-0.023	0.075	0.052	0.042	0.028
	(0.039)	(0.037)	(0.039)	(0.039)	(0.045)	(0.048)	(0.040)	(0.037)
AC: Data	0.416**	0.139	0.243	0.211	0.327*	0.471*	0.326*	0.090
	(0.155)	(0.140)	(0.144)	(0.143)	(0.167)	(0.186)	(0.143)	(0.148)
AC: Context	-0.090	-0.068	0.039	-0.060	-0.054	-0.013	-0.336**	-0.128
	(0.128)	(0.114)	(0.126)	(0.119)	(0.140)	(0.149)	(0.125)	(0.129)
AC: Scope	-0.010	-0.111	-0.104	-0.127	-0.070	0.119	0.110	0.145
	(0.109)	(0.111)	(0.111)	(0.109)	(0.124)	(0.133)	(0.115)	(0.108)
AC: Consequence	-0.089	-0.121	-0.130	-0.206	-0.183	-0.175	0.197	0.068
	(0.166)	(0.202)	(0.185)	(0.186)	(0.213)	(0.205)	(0.165)	(0.179)
N	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 12: Ordered probit regressions for the Co-Responsibility experiment

	Wrongful	Harmful	Serious	Harsh	Pot. Harm	Sensitive	Respons.	Clever
Not Patched	0.133	0.102	0.157	0.074	0.087	-0.370*	0.423***	-0.184
	(0.136)	(0.136)	(0.133)	(0.132)	(0.151)	(0.164)	(0.128)	(0.136)
Female	0.197	0.192	0.144	0.060	0.154	-0.045	0.151	0.029
	(0.147)	(0.153)	(0.151)	(0.143)	(0.162)	(0.175)	(0.142)	(0.149)
US birth	0.225	-0.758*	0.366	0.227	-0.337	-0.593	-0.509	0.214
	(0.356)	(0.298)	(0.274)	(0.234)	(0.433)	(0.456)	(0.356)	(0.391)
CFIP score	0.576***	0.391**	0.557***	0.385***	0.701***	1.087***	0.249*	0.364**
	(0.120)	(0.130)	(0.117)	(0.113)	(0.137)	(0.141)	(0.113)	(0.125)
Freq. aff by cybercrime	-0.007	0.035	-0.026	-0.048	0.034	0.011	0.097	0.002
	(0.084)	(0.089)	(0.104)	(0.094)	(0.107)	(0.118)	(0.100)	(0.098)
Fake personal info	-0.016	-0.154*	0.001	-0.121	0.025	-0.004	-0.051	0.066
	(0.059)	(0.065)	(0.066)	(0.063)	(0.070)	(0.066)	(0.062)	(0.070)
Media awareness	0.030	0.113*	0.093	0.076	0.041	-0.069	0.177**	0.064
	(0.048)	(0.050)	(0.048)	(0.048)	(0.056)	(0.062)	(0.054)	(0.054)
AC: Data	-0.271	-0.305	-0.343	-0.060	-0.071	-0.202	-0.197	0.185
	(0.206)	(0.219)	(0.189)	(0.218)	(0.206)	(0.270)	(0.202)	(0.201)
AC: Context	-0.359*	-0.359*	-0.286	-0.234	-0.553**	-0.251	-0.144	0.075
	(0.162)	(0.150)	(0.161)	(0.148)	(0.178)	(0.192)	(0.162)	(0.169)
AC: Scope	0.007	0.234	0.271	0.082	0.494**	0.384*	0.226	0.200
	(0.189)	(0.177)	(0.171)	(0.160)	(0.192)	(0.177)	(0.181)	(0.169)
AC: Patched	-0.294	-0.333	-0.184	-0.283	-0.277	-0.462	-0.359	0.032
	(0.234)	(0.229)	(0.209)	(0.212)	(0.286)	(0.284)	(0.225)	(0.240)
N	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 13: Ordered probit regressions for the Context experiment (vs. Bank)

	Wrongful	Harmful	Serious	Harsh	Pot. Harm	Sensitive	Respons.	Clever
Government	-0.055	0.013	-0.027	-0.030	0.147	-0.121	0.152	-0.023
	(0.119)	(0.121)	(0.125)	(0.116)	(0.139)	(0.142)	(0.118)	(0.116)
Non-Profit	0.048	-0.029	-0.222	0.030	0.099	-0.208	-0.361**	-0.185
	(0.123)	(0.124)	(0.122)	(0.121)	(0.140)	(0.155)	(0.120)	(0.121)
Org. size	0.055	0.064	0.045	0.053	0.133**	0.148**	0.059	0.142**
	(0.044)	(0.041)	(0.043)	(0.043)	(0.048)	(0.050)	(0.046)	(0.046)
Female	0.002	0.000	-0.044	-0.018	0.090	0.068	0.157	0.127
	(0.102)	(0.101)	(0.100)	(0.099)	(0.116)	(0.118)	(0.096)	(0.100)
US birth	-0.069	-0.094	0.071	-0.157	-0.292	0.158	0.116	-0.050
	(0.281)	(0.276)	(0.226)	(0.250)	(0.284)	(0.376)	(0.301)	(0.276)
CFIP score	0.354***	0.191*	0.376***	0.207**	0.405***	0.518***	0.139	0.135
	(0.073)	(0.077)	(0.073)	(0.080)	(0.085)	(0.077)	(0.075)	(0.073)
Freq. aff by cybercrime	-0.020	-0.027	-0.052	-0.044	-0.118	0.004	-0.026	0.047
	(0.064)	(0.063)	(0.064)	(0.060)	(0.073)	(0.077)	(0.064)	(0.069)
Fake personal info	-0.021	0.003	-0.016	-0.007	0.053	-0.078	0.013	-0.009
	(0.041)	(0.040)	(0.040)	(0.037)	(0.045)	(0.044)	(0.042)	(0.040)
Media awareness	-0.026	-0.046	0.030	-0.010	-0.030	-0.010	0.065	0.062
	(0.036)	(0.038)	(0.036)	(0.035)	(0.044)	(0.043)	(0.037)	(0.037)
AC: Data	0.023	0.019	0.029	0.017	0.376*	0.372*	-0.184	-0.053
	(0.153)	(0.151)	(0.153)	(0.152)	(0.161)	(0.156)	(0.123)	(0.142)
AC: Context	-0.003	0.066	-0.196	0.010	-0.024	-0.036	-0.101	-0.153
	(0.129)	(0.133)	(0.128)	(0.134)	(0.144)	(0.160)	(0.123)	(0.129)
AC: Scope	-0.152	0.051	-0.101	-0.035	-0.028	0.168	-0.022	0.035
	(0.122)	(0.136)	(0.126)	(0.123)	(0.144)	(0.142)	(0.127)	(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

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 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

To explore the extent to which respondents agree in their perceptions, we ran individual regressions for each of the 223 respondents in our experiment. The statistical power in the individual-level regressions is limited by the fact that each respondent rated only 25 vignettes.[131] 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 14 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(Sentence), which was significant at $p < 0.05$ in 34% of individual regressions. All the variables of interest except those involving organization type were significant at rates higher than the corresponding p level (i.e., the coefficient was significant at $p < 0.05$ for more than 5% of responses). Table 15 shows summary statistics for the coefficients across individual-level regressions. Figure 2 and Figure 3 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.

[131] As we discuss in Section 2, *supra*, we scaled back to 25 vignettes per respondent after a pilot study with 40 vignettes per person exhibited technical problems and high dropout rates.

Table 14: 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 15: 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 2: Distribution of β values over respondent-level models for non-log-scaled variables of interest

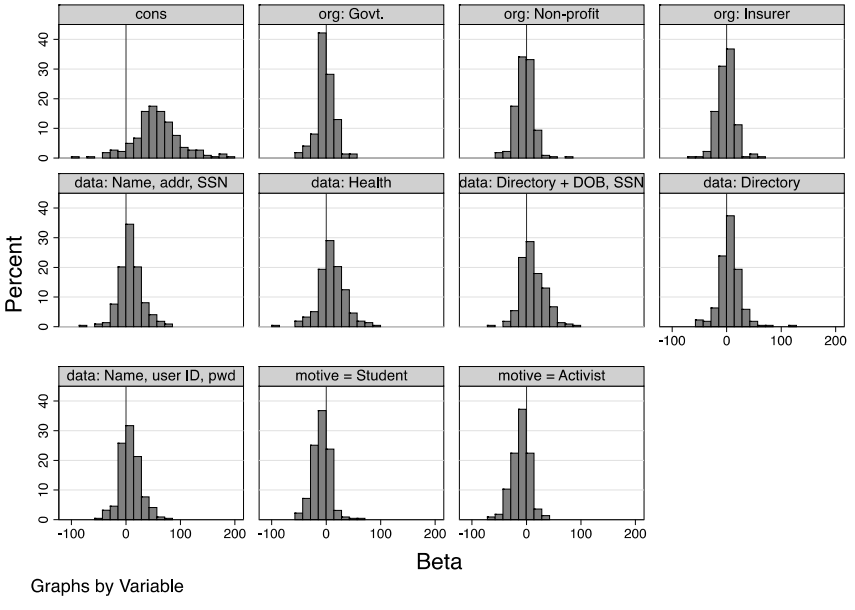

Figure 3: Distribution of β values over respondent-level models for log-scaled variables of interest