WHY CALIFORNIA ENACTED THE CCPA: ARE CONCERNS ABOUT DATA PRIVACY ENOUGH TO DRIVE STATE OR FEDERAL LEGISLATION?

by
Casey Masamitsu

A capstone submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Science in Data Analytics and Policy.

Baltimore, Maryland
December 2022

Abstract

Data creation and usage is common practice for many corporations, governments, and consumers. Ample research suggests that Americans on both sides of the political spectrum seek data privacy protections, yet no robust federal data privacy law exists. However, California recently passed a thorough data privacy regulation called the California Consumer Protection Act (CCPA), strengthened by the California Privacy Rights Act (CPRA). This study first creates individual "data privacy scores" for a sample of Americans and uses the scores to conduct regression analysis. Next, further analysis determines that Californians do not hold differing views about who should be responsible for implementing data privacy protections. The analysis results demonstrate that Californians' data privacy concerns are no more heightened than residents in other states. The finding that Americans share similar views on data privacy is substantial, as understanding what drives politicians to create policy is necessary to steer activists in a more efficacious direction.

Table of Contents

1. Introduction

Americans widely contend that data privacy is a critical issue and that the United States government should play a significant role in addressing users' data privacy concerns. To point, 81% of American adults recently expressed concern about their data privacy and reported having "very little/no control of the data [businesses] collect."[1] Additional research suggests that data privacy concerns also transcend the hyper-partisanship that characterizes contemporary American society.[2] Indeed, Morning Consult found that 86% of Democrats and 81% of Republicans believe Congress should make data privacy a "top" or "important" priority in 2021.[3] Given the current White House's celebratory sentiments towards bipartisan accomplishments,[4] it is remarkable that the federal government has yet not delivered an organized, comprehensive response to most American adults' shared concern.

---

[1] Auxier, Brooke, Lee Rainie, Monica Anderson, Andrew Perrin, Madhu Kumar, and Erica Turner. "Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information." *Pew Research Center: Internet, Science & Tech* (blog), November 15, 2019. https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/.

[2] Dimock, Michael, and Richard Wike. "America Is Exceptional in Its Political Divide." Accessed December 10, 2022. https://pew.org/3bDV6Fa.

[3] Sam Sabin. "States Are Moving on Privacy Bills. Over 4 in 5 Voters Want Congress to Prioritize Protection of Online Data." *Morning Consult* (blog), April 27, 2021. https://morningconsult.com/2021/04/27/state-privacy-congress-priority-poll/.

[4] The White House. "President Biden's Bipartisan Infrastructure Law." Accessed December 10, 2022. https://www.whitehouse.gov/bipartisan-infrastructure-law/.

To date, only California, Colorado, Connecticut, Utah, and Virginia have passed comprehensive data privacy laws.[5] These five states' laws "have several provisions in common, such as the right to access and delete personal information and to opt-out of the sale of personal information, among others."[6] California was the first state to pursue this type of legislation with the California Consumer Protection Act (CCPA) and then the California Privacy Rights Act (CPRA).[7] Data privacy rights activists commonly view the CCPA and CPRA as the current "gold standard" and credit the legislation for inspiring similar iterations in Colorado, Connecticut, Utah, and Virginia.[8]

Considering California's influential role as a leader in protecting data privacy, understanding how and why lawmakers passed the CCPA and CPRA provides valuable tools for legislators, activists, and American voters. To distill any meaningful insights, though, it is crucial to know if Californian voters—who adopted the state's data privacy enhancement (CPRA) with 56% of the vote[9]—have unique views on data privacy.

---

[5] Staszkiw, Michael, and Anna Mercado Clark. "Navigating Different Obligations of State Privacy Statutes for 2023." *Rochester Business Journal* 38, no. 21 (October 21, 2022): 44.

[6] National Conference of State Legislatures. "State Laws Related to Digital Privacy." Accessed December 10, 2022. https://www.ncsl.org/research/telecommunications-and-information-technology/state-laws-related-to-internet-privacy.aspx.

[7] Bonta, Rob. "California Consumer Privacy Act (CCPA)." State of California - Department of Justice - Office of the Attorney General, October 15, 2018. https://oag.ca.gov/privacy/ccpa.

[8] Voss, W Gregory. "THE CCPA AND THE GDPR ARE NOT THE SAME: WHY YOU SHOULD UNDERSTAND BOTH." *Competition Policy International*, 2021, 8.

[9] Padilla, Alex. "Statement of Vote: General Election November 3, 2020." Secretary of State of California, November 3, 2020.

Therefore, this research aims to answer the question: Do concerns about data privacy protections and which entity should bear the responsibility to enact them vary between Californians and the rest of America?

To begin answering this question, this study uses feature engineering and regression analyses to first investigate if Californians have heightened concerns about data privacy when compared with the rest of America. Next, further regression is conducted to examine which entities Californians believe should bear responsibility for data privacy protection regulations. The findings support the hypothesis that Californians do not hold significantly different views than Americans outside of California. However, additional analysis suggests that other demographic independent variables, such as age and race, play a part in one's views on data privacy and ideas about which party should claim responsibility for data privacy protection.

2. Literature Review and Theoretical Framework

2.1 Social & Governmental Context

In 2000, the Federal Communications Commission (FCC) called on Congress to establish a basic consumer privacy law to accommodate the rapid growth in the online consumer marketplace.[10] The federal government has still not passed such a bill, yet data has continued growing exponentially.[11] The International Data Corporation (IDC)—a

[10] Pitofsky, Robert, Sheila F Anthony, Mozelle W Thompson, Orson Swindle, and Thomas B Leary. "Federal Trade Commission." *Privacy Outline*, n.d., 208.

[11] Woodie, Alex. "Big Growth Forecasted for Big Data." Datanami, January 12, 2022. https://www.datanami.com/2022/01/11/big-growth-forecasted-for-big-data/.

global market intelligence firm—noted 64.1 zettabytes of data created in 2020 and projected that "[t]he amount of digital data created over the next five years will be greater than twice the amount of data created since the advent of digital storage."[12] While IDC's report represents data writ large, consumers have expressed pointed concerns regarding a subset of their data known as "personally identifiable information" (PII). The U.S. Department of Labor defines PII as "any representation of information that permits the identity of an individual to whom the information applies to be reasonably inferred by either direct or indirect means."[13]

There is broad consensus about these definitions. However, elected officials, corporations, and Americans have differing opinions about and approaches to protecting users' data privacy. In fact, businesses may legally use personally identifiable information to target end-users and strategically bolster sales or marketing efforts, prioritizing capitalist needs over personal ones.[14] Worse, current United States law has no comprehensive provisions protecting users' data from corporate employees or potential hacking efforts. Peter Zatko, the former security chief at Twitter, provides evidence to support this claim. In his recent testimony to the United States Senate, Mr. Zatko stated: "It's not far-fetched to say that employees inside [Twitter] could take over the accounts

---

[12] Woodie, Alex. "Big Growth Forecasted for Big Data."

[13] U.S. Department of Labor. "Guidance on the Protection of Personal Identifiable Information." Accessed September 19, 2022. https://www.dol.gov/general/ppii.

[14] Godinho de Matos, Miguel. "Consumer Consent and Firm Targeting After GDPR: The Case of a Large Telecom Provider." *Management Science* 68, no. 5 (2021): 3330–78.

of all of the senators in this room."[15] Zatko spoke from experience, as hackers seized several prominent users' Twitter accounts in 2020 and successfully received $121,000 in Bitcoin payments from the hacked users' followers on the social media platform.[16]

2.2 California's Role and Prominence

Acknowledging the potential harm associated with unregulated personal data and technological security, it is, again, noteworthy that there has been no organized federal response in this area. Instead, the federal government has a pastiche of sector-specific data privacy regulations,[17] which constituents and elected leaders in at least three states have deemed insufficient.[18] Unlike the federal government, these states acted swiftly, introducing and passing comprehensive data privacy laws intended to enhance consumers' data security over the last four years.[19] California was the first state to successfully legislate this issue with the California Consumer Protection Act (CCPA).

---

[15] Rebecca Kern and Eric Geller. "Twitter Whistleblower to Congress: Your Data Is at Risk Too." POLITICO. Accessed September 19, 2022. https://www.politico.com/news/2022/09/13/whistleblower-zatko-testimony-agrawal-twitter-00056291.

[16] Leswing, Kif. "Twitter Hackers Who Targeted Elon Musk and Others Received $121,000 in Bitcoin, Analysis Shows." CNBC. Accessed September 24, 2022. https://www.cnbc.com/2020/07/16/twitter-hackers-made-121000-in-bitcoin-analysis-shows.html.

[17] Thorin Klosowski. "The State of Consumer Data Privacy Laws in the US (And Why It Matters)." Reviews for the Real World, September 6, 2021. https://www.nytimes.com/wirecutter/blog/state-of-privacy-laws-in-us/.

[18] Thorin Klosowski. "The State of Consumer Data Privacy Laws in the US (And Why It Matters)."

[19] Thorin Klosowski. "The State of Consumer Data Privacy Laws in the US (And Why It Matters)."

Passed in 2018, the CCPA took effect on January 1, 2020, and protects California's data consumers and generators by providing the following assurances:

> The right to know about the personal information a business collects about them and how it is used and shared; The right to delete personal information collected from them (with some exceptions); The right to opt-out of the sale of their personal information; and The right to non-discrimination for exercising their CCPA rights.[20]

Each business must comply with these stipulations when it has gross annual revenue of over $25 million; buys, receives, or sells the personal information of 50,000 or more California residents, households, or devices; or if it derives 50% or more of its annual revenue from selling California residents' personal information.[21] Importantly, under the CCPA, a business only has to comply with the law when it qualifies under one of these criteria, keeping businesses and consumers in mind.[22]

Nevertheless, California again led the country in advancing data privacy protections in 2020 when Californians voted to adopt the California Privacy Rights Act (CPRA). The CPRA, informally referred to as "CCPA 2.0,"[23] goes into full effect on January 1, 2023, and significantly extends the protections outlined in the CCPA. For instance, in addition to the Attorney General's enforcement powers detailed in the CCPA, the CPRA created the California Privacy Protection Agency. This agency houses a "five-

---

[20] Bonta, Rob. "California Consumer Privacy Act (CCPA)."

[21] Bonta, Rob. "California Consumer Privacy Act (CCPA)."

[22] Bonta, Rob. "California Consumer Privacy Act (CCPA)."

[23] Bloomberg Law. "What's the Difference Between CCPA & CPRA." Accessed December 10, 2022. https://pro.bloomberglaw.com/brief/the-far-reaching-implications-of-the-california-consumer-privacy-act-ccpa/.

member board that consists of experts in privacy, technology, and consumer rights"[24] who are wholly dedicated to preserving consumers' data privacy rights. Since this development, California has become a potential model for other states and is changing the way companies do business."[25]

2.3 Potential Impacts: Policymakers v. Constituents

Noting Americans' nonpartisan desire for government officials to pass data privacy concerns, the role of political actors also becomes relevant. Political actors at the state and federal levels introduce, co-sponsor, and vote for or against legislation while presumably representing their constituents' interests.[26] However, existing regulations do not reflect the public's opinion across the country.[27] This literature search did not yield any comparative analysis to assess the differences between California residents and the rest of the United States. Rather than a comparative analysis between states, current research includes public opinion analyses that quantify national consumers' opinions about data privacy within specific industries or political parties. As a result, it is currently unknown if California residents have stronger opinions about their data privacy than residents of other states, which may have influenced California to act first.

---

[24] California, State of. "Frequently Asked Questions (FAQs) - California Privacy Protection Agency (CPPA)." Accessed December 10, 2022. https://cppa.ca.gov/faq.html.

[25] Bloomberg Law. "What's the Difference Between CCPA & CPRA."

[26] Go Government. "Working in the Legislative Branch." Accessed September 24, 2022. https://gogovernment.org/all-about-government-jobs/working-in-the-legislative-branch/.

[27] Auxier, Brooke, Lee Rainie, Monica Anderson, Andrew Perrin, Madhu Kumar, and Erica Turner. "Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information." *Pew Research Center*

Still, a national study conducted by the Public Library of Science (PLOS) found that partisan self-interest is an important driver of people's support for regulating targeted political advertising.[28] In PLOS's study, the researchers found Republican voters more likely to oppose data privacy regulation if first informed that Republicans benefitted historically from targeted political advertising in elections.[29] Even so, while partisan self-interest may inform views on data privacy protections, a survey in 2021 conducted by Privacy for America found that 92% of United States adults, regardless of political affiliation, report it is "very or somewhat important that Congress pass legislation to protect consumer data privacy."[30]

2.4 Additional Considerations

The healthcare industry also provides some informative context because data privacy concerns within healthcare have been heavily researched, especially since COVID-19. Pew Research found that 91% of U.S. adults are concerned that companies might abuse their healthcare data.[31] While there is not necessarily a direct correlation

[28] Baum, Katharina, Stefan Meissner, and Hanna Krasnova. "Partisan Self-Interest Is an Important Driver for People's Support for the Regulation of Targeted Political Advertising." *PLOS ONE* 16, no. 5 (May 12, 2021): e0250506. https://doi.org/10.1371/journal.pone.0250506.

[29] Baum, Katharina, Stefan Meissner, and Hanna Krasnova. "Partisan Self-Interest Is an Important Driver for People's Support for the Regulation of Targeted Political Advertising."

[30] Privacy for America. "Nationwide Opinion Research on Data Privacy." Accessed September 22, 2022. https://www.privacyforamerica.com/nationwide-opinion-research-on-data-privacy-pdf/.

[31] Auxier, Brooke, Lee Rainie, Monica Anderson, Andrew Perrin, Madhu Kumar, and Erica Turner. "Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information." *Pew Research Center*

between users' views on healthcare data and their other personally identifiable information, Pew's results suggest that the national surveys might provide baseline figures for Americans' views on data privacy. If asked about sector-specific or business-specific concerns, Americans' concerns about data privacy might be even more significant, as is the case with healthcare.

Comprehensive data privacy laws are taking effect over the next year in several states, including California.[32] While other states have attempted to follow in California's footsteps, few have successfully passed such legislation. This research, then, explores Californians' views on data privacy versus the rest of the United States. Understanding how or if data privacy concerns vary by the state is valuable information for the states and federal government. By isolating which state's residents care the most about data privacy regulation, states or the federal government can effectively prioritize its agenda. This research will look at the state of residence, race, political ideology, age, educational attainment, and other demographic variables to determine if Californians' relevant attributes contributed to California's passing of the CCPA or if politicians catalyzed this policy shift. Regardless, this study aims to explore Californians' collective influence on these policies, which may provide a viable model for other states interested in passing similar regulations.

3. Data and Methods

[32] Thorin Klosowski. "The State of Consumer Data Privacy Laws in the US (And Why It Matters)."

3.1 Data

This study relies on an observational, cross-sectional dataset containing survey results to assess data privacy concerns between California and the rest of the United States. Pollsters with The Associated Press-NORC Center for Public Affairs conducted this survey via the web and telephone communications between April 11, 2018, and April 16, 2018. The Associated Press and NORC at the University of Chicago funded the project. The resulting dataset includes survey question responses and demographics of 1,140 participants. Each observation represents a United States adult aged 18 or over. The analysis excludes 338 participants who skipped or refused to answer at least one question. As such, the resulting analyses include data from 802 participants who answered each question that informed this research.

The NORC survey presented the interviewees with questions related to three distinct areas: foreign relations, teacher pay and protests, and attitudes toward privacy and data security. This analysis focuses on the data acquired in the privacy and data security section, aptly coded as "PRIV" (see Appendix for the complete list of these questions). Several other demographics constitute the independent variables in the multivariate analysis; these demographics include political party, urban versus rural residence, marital status, age, employment status, educational attainment, race, household income, gender, and California residents versus non-Californian residents. Dichotomized into binary variables, each independent variable represents a distinct group of individuals.

Participants' answers to 14 of the interview questions facilitated the creation of five dependent variables. Interviewers recorded participants' responses on a one-to-five Likert scale. Employing data transformation, participants' responses in this study range

between 0-100. In this context, a lower value indicates the respondent has a lower data privacy tolerance score or believes the entity in question should bear lesser responsibility. A higher value reflects a respondent's heightened concerns regarding data privacy protections or a stronger belief that the entity proposed should bear data privacy protection responsibility. The transformation of these responses to a discrete range improves readability and eases the interpretation of the results. Below, Tables 1 and 2 depict the variable definitions (including how to interpret the transformed ranges for the dependent variables) and summary statistics, respectively.

Table 1. Variable Definitions

Variables	**Description**	**Definition**
californian	State of residence	1 = Californian 0 = Other
party	Political party	1 = Democrat 0 = Other
urban	Lives in urban setting	1 = Urban 0 = Other
married	Marital status	1 = Married 0 = Other
over40years	Age	1 = >40 Years Old 0 = < 40 Years Old
employed	Employment status	1 = Employed 0 = Not employed
graduate	Highest educational attainment	1 = Bachelor's degree or higher 0 = Less than Bachelor's degree
white	Race	1 = White 0 = Non-white
over50k	Household income	1 = Over $50,000 0 = Under $50,000
male	Gender	1 = Male-identifying

		0 = Not male-identifying
users	Respondent's opinion regarding how much onus lies on the individual user regarding data privacy protections	Discrete ranges between 0-100: > 87.5 = Very Large 62.5-87.49 = Large 37.5-62.49 = Moderate 12.5-37.49 = Small < 12.49 = None at all
government	Respondent's opinion regarding how much onus lies on the federal government regarding data privacy protections	Discrete ranges between 0-100: > 87.5 = Very Large 62.5-87.49 = Large 37.5-62.49 = Moderate 12.5-37.49 = Small < 12.49 = None at all
socialmedia	Respondent's opinion regarding how much onus lies on social media companies regarding data privacy protections	Discrete ranges between 0-100: > 87.5 = Very Large 62.5-87.49 = Large 37.5-62.49 = Moderate 12.5-37.49 = Small < 12.49 = None at all
thirdparties	Respondent's opinion regarding how much onus lies on third parties regarding data privacy protections	Discrete ranges between 0-100: > 87.5 = Very Large 62.5-87.49 = Large 37.5-62.49 = Moderate 12.5-37.49 = Small < 12.49 = None at all
score	The feature-engineered data privacy tolerance score	Discrete ranges between 0-100: > 87.5 = Extremely concerned 62.5-87.49 = Very concerned 37.5-62.49 = Moderately concerned 12.5-37.49 = Not very concerned < 12.49 = Not at all concerned

Table 2. Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Median	Max
californian	802	0.10	0.30	0	0	1
party	802	0.38	0.49	0	0	1
urban	802	0.32	0.47	0	0	1
married	802	0.51	0.50	0	1	1
over40years	802	0.56	0.50	0	1	1
employed	802	0.64	0.48	0	1	1
graduate	802	0.35	0.48	0	0	1
white	802	0.64	0.48	0	1	1
over50k	802	0.55	0.50	0	1	1
male	802	0.40	0.49	0	0	1
users	802	79.68	23.73	0	75	100
government	802	62.22	31.03	0	75	100
socialmedia	802	86.41	21.51	0	100	100
thirdparties	802	67.43	30.15	0	75	100
score	802	66.77	21.00	7.25	67.75	100.00

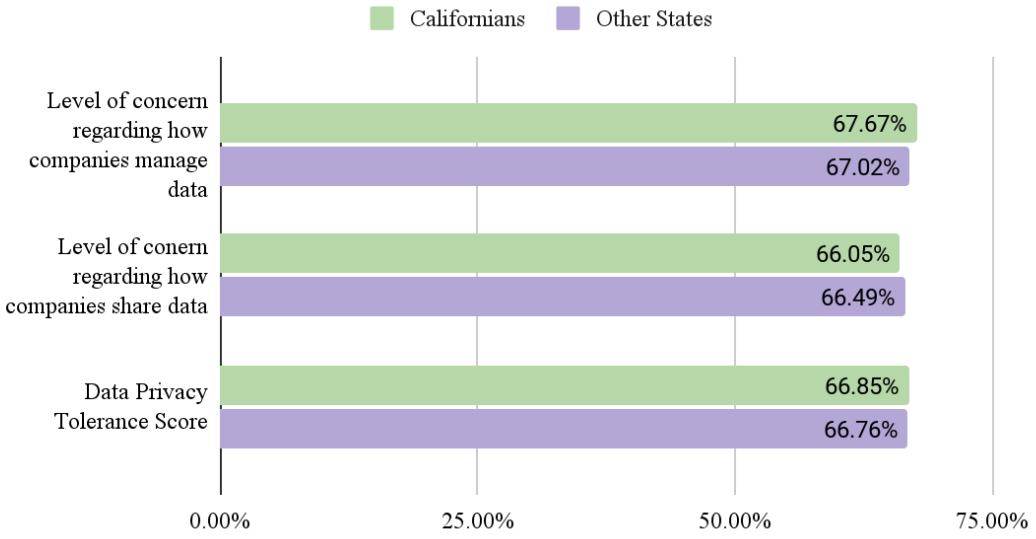
**Several states included only a handful of respondents. As such, a "californian" independent variable was created using the respondents' state of residence instead of utilizing fixed-effects regression.*

3.2 Data Privacy Tolerance Score

The first dependent variable (score) is a feature-engineered data privacy tolerance score. Two umbrella questions and corresponding hypotheticals informed the feature engineering process. The first question measured respondents' broad concerns regarding how companies manage their data in four individual contexts, eliciting four unique responses. The second question asked respondents to provide their concerns about how social media companies choose to access and share their data with four types of external entities. Similar to question one, the second umbrella question resulted in four unique responses per participant. Surveyors utilized a Likert scale to measure each participant's eight total responses; each of which is reflected in a range from 0-100 (Table 1).

The analysis used RStudio software to average the participants' responses from each umbrella question and subsequently average the two together to create the data privacy score (see Appendix for code). The resulting data privacy score ranges from 0-100, where a score of 100 indicates the respondent answered “very concerned” to every question, and a score of 0 indicates the respondent answered “not at all concerned” to each question. Figure 1 displays the average answer to each umbrella question and the average data privacy scores for Californians versus respondents in other states.

Figure 1. Question Result Means & Data Privacy Tolerance Score



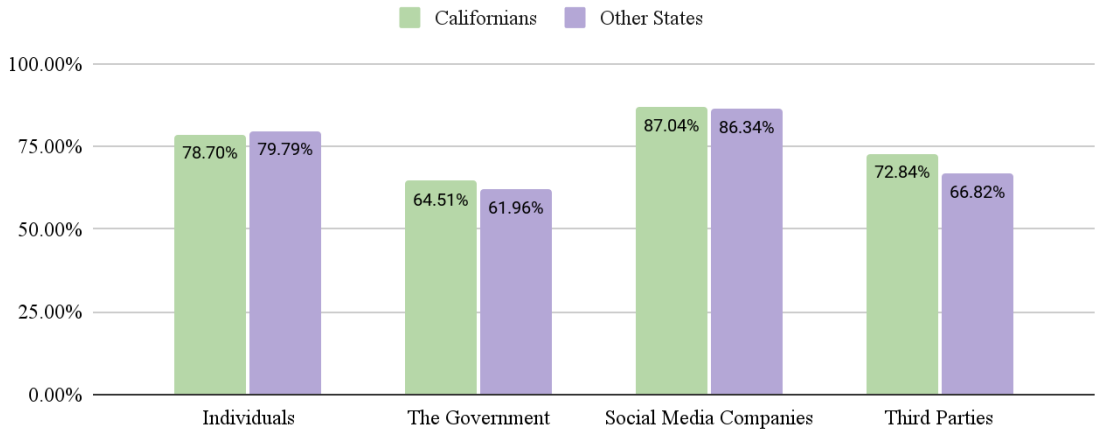
**Additional details describing the feature engineering process are included in Appendix.*

3.3 Additional Dependent Variables

Additional considerations included participants’ views regarding if individual users, the federal government, social media companies, or third parties should claim responsibility for data privacy protections. These views on data privacy protection responsibility comprise this analysis's four remaining dependent variables. Like the data

privacy tolerance score, surveyors used a Likert scale of one-to-five to gauge participants' opinions on which entity should be responsible for instituting data privacy protections. Also transformed to a range of 0-100, a higher score indicates a more decisive view that the entity communicated by the surveyor should bear a greater responsibility for ensuring data privacy protections. Table 1 and Figure 2 also include guidance on interpreting the value of these four dependent variables.

Figure 2. How Strongly Respondents Feel Each Respective Entity Should Be Responsible for Data Privacy Protections



None (>12.5%), Small (12.5-37.49%), Moderate (37.5-62.49%), Large (62.5-87.49%), Very Large (>87.5%)

3.4 Multivariate Regression

The analysis continued with multivariate regression. The first model displays the regression between the data privacy tolerance score (score) and the independent variables: californian, party, urban, married, over40years, employed, graduate, white, over50k, and male. The multivariate regression also considered the relationships between the same independent variables and the four dependent variables – users, government,

socialmedia, and thirdparties. Ultimately, the multivariate regressions yielded five separate models discussed in the following section.

4. Results

4.1 Data Privacy Tolerance Score Regressions

Table 3 depicts bivariate and multivariate regression results for the first dependent variable: the data privacy tolerance score. The first is a bivariate model, which displays the sole impact of residing in California on the data privacy tolerance score. The results of the model imply a 0.09% increase in an individual's data privacy tolerance score if that participant resides in California. Put differently, Californians' average data privacy tolerance score is 0.09% higher than non-Californians (also depicted in Figure 1). However, researchers and policymakers would have difficulty distilling meaningful conclusions from this model because the coefficient is not statistically significant at any confidence level. The bivariate model, then, suffers from omitted variable bias, further evidenced by the adjusted R^2 of -0.001.

Table 3. Data Tolerance Privacy Score

	Dependent variable:	
	score	
	Bivariate	Multivariate
	(1)	(2)
californian	0.09 (2.41)	-1.03 (2.40)
party		-0.16 (1.59)
urban		-0.30 (1.66)
married		-1.41 (1.62)
over40years		4.40*** (1.52)
employed		-1.47 (1.62)
graduate		-0.50 (1.60)
white		-9.18*** (1.65)
over50k		1.37 (1.66)
male		2.22 (1.52)
Constant	66.76*** (0.78)	70.64*** (2.28)
Observations	802	802
R^2	0.0000	0.05
Adjusted R^2	-0.001	0.04
Residual Std. Error	21.01 (df = 800)	20.55 (df = 791)
F Statistic	0.001 (df = 1; 800)	4.55*** (df = 10; 791)
Note:		*p<0.1; **p<0.05; ***p<0.01

Building on the bivariate model, multivariate regression created the same model using several additional independent variables: political affiliation, urban versus rural residency, marital status, age, employment status, educational attainment, race, household income, and gender (described in Table 1). After including these control variables, the coefficient for residing in California decreases to -1.03. The coefficient of -1.03 means, when controlling for the above demographic independent variables, residing in California is associated with a 1.03% lower data privacy score. However, this coefficient remains insignificant at all confidence levels, further suggesting that Californians' data privacy tolerance scores do not vary significantly from those in other states.

Several other variables in the multivariate regression model did not have statistical significance in this analysis. These variables included party affiliation, urban residents, marital status, employment status, educational attainment, household income, and gender. The lack of statistical significance for these independent variables implies the variables have a little-to-no respective impact on an individual's data privacy tolerance score. Consequently, the results of the regression are supportive of the hypothesis that residing in California does not have a significant impact on one's data privacy tolerance score.

Notwithstanding, the multivariate model produced two statistically significant variables: individuals over 40 years of age and white participants. With a p-value of less than 0.01, each of the resulting coefficients is statistically significant at the 99% confidence level. The constant in this model is 70.64%, squarely in the "very concerned" category. Therefore, analysts and policymakers can associate individuals over 40 years old with a 4.40% increase in their respective data privacy tolerance score (up to 75.04% from the constant), suggesting a generational interest that could potentially rise.

In contrast, analysts can associate being white with a -9.18% decrease in their respective data privacy tolerance score. Such a large coefficient reduces the average data privacy tolerance score for white respondents to 61.46%, descending the average score into the "moderately concerned" category. The latter score undergirds a relationship between personal agency and sociopolitical trust that lawmakers could leverage to their respective advantage.

The low adjusted R^2 value present in each model indicates that both models still suffer from omitted variable bias. Notwithstanding, each model suggests residing in

California does not significantly impact one's data privacy tolerance score. This outcome again supports the hypothesis that Californian residents do not have more significant concerns regarding data privacy protections, though other demographic variables might.

4.2 Responsible Parties Regression

Because the data privacy tolerance scores for Californian residents do not vary from those living in other states, this research next analyzes if California residents hold unique opinions about who should bear the responsibility for implementing data privacy protections. The understanding of perceived culpability could assist stakeholders when exploring the contributing factors Californian lawmakers considered when enacting the CCPA or CPRA. Accordingly, this analysis includes the creation of four multivariate regression models, each depicted in Table 4. While it is worth noting that all of the respondents, on average, believe all four of the entities offered in the survey should have at least a "large" level of involvement, not all demographics share a view on who should bear the greatest responsibility or how to quantify said responsibility.

Table 4. Responsibility for Data Privacy Protections

	Dependent variable:			
	users	government	socialmedia	thirdparties
	(1)	(2)	(3)	(4)
californian	-1.22 (2.79)	0.30 (3.29)	0.58 (2.32)	5.02 (3.10)
party	-1.22 (1.81)	5.49** (2.21)	1.79 (1.65)	2.41 (2.29)
urban	3.02* (1.82)	4.97** (2.29)	2.97* (1.69)	0.88 (2.33)
married	0.55 (1.81)	2.80 (2.40)	0.96 (1.74)	2.55 (2.32)
over40years	3.98** (1.78)	9.40*** (2.15)	5.69*** (1.57)	12.53*** (2.20)
employed	4.40** (1.83)	-5.99*** (2.31)	2.09 (1.66)	-1.99 (2.29)
graduate	-0.26 (1.80)	-6.18*** (2.28)	1.34 (1.62)	0.07 (2.26)
white	-2.22 (1.90)	-10.97*** (2.32)	1.75 (1.74)	-2.99 (2.27)
over50k	1.88 (1.86)	-0.75 (2.50)	1.64 (1.74)	0.87 (2.42)
male	-0.27 (1.77)	-1.44 (2.25)	1.20 (1.63)	-2.26 (2.24)
Constant	74.56*** (2.68)	65.88*** (3.21)	76.76*** (2.54)	61.03*** (3.08)
Observations	802	802	802	802
R^2	0.02	0.10	0.03	0.06
Adjusted R^2	0.01	0.08	0.02	0.04
Residual Std. Error (df = 791)	23.64	29.70	21.28	29.49
F Statistic (df = 10; 791)	1.62*	8.34***	2.76***	4.67***
Note:				*p<0.1; **p<0.05; ***p<0.01

4.2.1 Individual Users

The first model depicts respondents' perceived onus that individual users expressed about protecting their data. The results of this model display that, on average, participants in the survey concede that individual users should have at least a "large" part in protecting their own data. The coefficient for residents of California is -1.22, which indicates that Californians believe the data privacy responsibilities should land on the individual user slightly lower than the rest of Americans. However, this coefficient is not statistically significant at any confidence level. A few groups reported that individual users should bear more responsibility in this regard – those living in an urban environment, those over 40 years old, and employed individuals. Still, each respective

increase is less than 5%, which shares the average opinion that individual users should have a “large” responsibility in protecting themselves.

4.2.2 The Federal Government

The second regression model in Table 4 reveals how strongly the respondents believe that the federal government should regulate data privacy protections. Six of the ten independent variables in this model are statistically significant, suggesting a divisive outlook on views of the federal government’s responsibility for data privacy protections. Further, of the four proposed entities, the perceived onus of the federal government is the lowest at 64.51% for Californians and 61.96% for non-Californians (Figure 2).

In the multivariate regression, three demographics believe the federal government should bear more responsibility than the rest of the participants – Democrats, respondents living in an urban setting, and respondents over 40 years old. Compared with their respective counterparts, Democrats’ responses were 5.49% higher, those in an urban environment were 4.97% higher, and those over 40 were 9.40% higher. These three coefficients are all statistically significant. With a constant of 65.88%, the average response for Democrats over 40 who live in an urban environment jumps to 85.74% – near the “very large” category.

In contrast, three demographics responded lower than the rest of the participants – white respondents, those with a Bachelor’s degree, and employed individuals. Meaning, with a constant of 65.88%, the average response for employed white respondents with at least a Bachelor’s degree is only 42.74% – less than half the perceived onus from Democrats over 40 in urban environments. A response of 42.74% lowers the average

response from the “large responsibility” category to the lower end of the“moderate responsibility” category.

The third model also resulted in the highest adjusted R^2 value of any of the multivariate models created, 0.08. An adjusted R^2 of 0.08 is relatively high compared to the other models. The adjusted R^2 suggests that the demographics included in this model explain 8% of the change in the dependent variable – the federal government's perceived data privacy protection onus.

4.2.3 Social Media Companies

The third model demonstrates respondents’ perceived responsibility surrounding data privacy protection of social media companies. Notably, on average, respondents believe that social media companies should bear the most responsibility (compared to the other three entities included in this study) in ensuring that their users’ data is protected. Californians responded at 87.04% and other Americans at 86.34% – each response nearly entering the “very large responsibility” category. The coefficient for residing in California is only 0.58, meaning we can associate living in California with only a 0.58% increase in the average response. However, few of the variables included in the model are statistically significant, including residing in California. In addition, the adjusted R^2 is only 0.02 – suggesting that variables included in this model have little-to-no impact on one’s perception of social media companies' data privacy responsibilities.

4.2.4 Third Parties

The highest coefficient for Californians is present in the fourth model regarding third parties. This coefficient is 5.02; therefore, the model states lawmakers can associate being from California with a 5.02% increase in the average response for their perceived responsibility of third parties. Indeed, as shown in Figure 2, Californians feel more decisively about third parties' responsibilities regarding data privacy protections, 72.84%, versus non-Californians, 66.82%.

Yet, the Californian coefficient in model four is not statistically significant at any confidence level, meaning further research is likely needed to make such a claim. In fact, the only demographic with a statistically significant coefficient is individuals over 40 years old. The coefficient for these individuals is 12.53, meaning that, on average, respondents over 40 years old responded 12.53% higher than those under 40. Still, this model suggests that residing in California, similarly to the other models, has no statistical significance in who the respondent deems responsible for data privacy protections.

4.2.5 Discussion

While Californians do not appear to have varying views on data privacy protection responsibility from non-Californians, age has a statistically significant impact in each of the four multivariate regression models. In fact, age is the only independent variable that is statistically significant across all five multivariate models. Compared to respondents under 40 years old, those over 40 believe that individual users should be held 3.98% more responsible, the federal government 9.40% more responsible, social media

companies 5.69% more responsible, and third parties 12.53% more responsible. On average, individuals over 40 responded that each of the four proposed entities should have at least a "large" responsibility in protecting their data privacy. Crucially, these coefficients are all statistically significant at the 99% confidence level, illustrating that people over 40 clearly have heightened concerns about data privacy and believe every proposed entity should claim more responsibility for consumers' data privacy protections.

Four independent variables showed no statistical significance about which party should be responsible for data privacy protections: residing in California, marital status, household income, and gender. The lack of statistical significance among these four independent variables implies that these demographics alone do not strongly affect one's views of liability in providing data privacy protections. Notably, the independent variable in focus – residing in California – has no statistical bearing in any of the six models created through regression. The lack of significance among all six models for Californians strongly supports the hypothesis that Californians do not have heightened data privacy concerns or unique views about which entity should be responsible for data privacy protections.

5. Conclusion

The findings of this research support the hypothesis that views around data privacy from those living in California are not statistically different from those living outside of California. In fact, Californians generally agree on the roles that individual

users, the federal government, social media companies, and third parties should have in ensuring that data privacy protections are enacted.

In a rare phenomenon, Americans' desires for data privacy protections appear nonpartisan, and, in an increasingly divided America, unity among this desire is especially unique. Assuming political actors represent their respective constituents' interests in good faith, the lack of action suggests other forces more directly influence their politicking. The CCPA and CPRA clearly reflect the interests of Californians. Yet, the vast majority of Americans' personally identifiable information remains abused, mishandled, or worse – vulnerable to hijacking or at a heightened risk of compromise. Lawmakers at the state and federal levels should interpret the results of this research as an unmistakable indication that Americans have coalesced around a common concern regarding data privacy. Given these factors, enacting a comprehensive law will likely not attract significant public opposition.

Still, the findings of this research leave a host of unanswered questions. The most obvious among them: If the American people are not a significant influence on the majority of state or federal responses to data privacy protections, what is/are? Future research could include an analysis of each state's response to data privacy protections, if any. Understanding which forces outside political actors' constituents' views impact or impede legislative decisions will prove pertinent to lawmakers and activists.

Additionally, since the state of Californian seemingly responded to their residents' data privacy protection desires, research into how California prioritized and enacted the CCPA and CPRA could prove worthwhile. Is the state government of California more productive than other states? Do Californian political agendas more

accurately reflect the highest priorities of their constituents? Alternatively, were lobbyists or activist groups directly involved in pushing for data privacy protections within California? What gave these interest groups such authority? Another possibility may be the presence of Silicon Valley within California – perhaps the increased occupancy of technology companies ignited such prioritization.

The analysis conducted in this research could also iteratively expand. Models could be recreated and informed by a different public opinion survey, specifically one with different data privacy questions or a survey more recent than 2018. Additionally, new models may benefit from additional independent control variables, especially socioeconomic data. Due to the statistical significance of one's age on data privacy views, age could be disseminated into age ranges to more specifically assess which age groups hold the highest data privacy concerns. Finally, the results of the second model (Section 4.2.2) displaying views of the federal government's involvement in data privacy protections could be built upon. Specifically, the question that informed this model inquired about the federal government, leaving views on state governments unfounded. Moreover, this model resulted in the most divisive responses, indicating a lack of alignment about the federal government's involvement in regulating data privacy protections.

Data privacy protections remain a high-priority issue for the majority of Americans. While there are slight to moderate deviations regarding how concerned individuals are about the current state of data privacy, harmony exists across party lines in the desire for comprehensive protection. Knowing this, the lack of prioritization by political actors remains a mystery. The creation of data and the associated risk of

vulnerability continues to grow exponentially – the federal government would be wise to respond to Americans' concerns sooner than later.

6. References

America, Privacy for. "Nationwide Opinion Research on Data Privacy." *Privacy for America* (blog). Accessed September 22, 2022. https://www.privacyforamerica.com/nationwide-opinion-research-on-data-privacy-pdf/.

Auxier, Brooke, Lee Rainie, Monica Anderson, Andrew Perrin, Madhu Kumar, and Erica Turner. "Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information." *Pew Research Center: Internet, Science & Tech* (blog), November 15, 2019. https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/.

Baum, Katharina, Stefan Meissner, and Hanna Krasnova. "Partisan Self-Interest Is an Important Driver for People's Support for the Regulation of Targeted Political Advertising." *PLOS ONE* 16, no. 5 (May 12, 2021): e0250506. https://doi.org/10.1371/journal.pone.0250506.

Bloomberg Law. "What's the Difference Between CCPA & CPRA." *Bloomberg Law* (blog). Accessed December 10, 2022. https://pro.bloomberglaw.com/brief/the-far-reaching-implications-of-the-california-consumer-privacy-act-ccpa/.

Bonta, Rob. "California Consumer Privacy Act (CCPA)." State of California - Department of Justice - Office of the Attorney General, October 15, 2018. https://oag.ca.gov/privacy/ccpa.

California, State of. “Frequently Asked Questions (FAQs) - California Privacy Protection Agency (CPPA).” Accessed December 10, 2022. https://cppa.ca.gov/faq.html.

Dimock, Michael, and Richard Wike. “America Is Exceptional in Its Political Divide.” Accessed December 10, 2022. https://pew.org/3bDV6Fa.

Duhigg, Charles. “How Companies Learn Your Secrets.” *The New York Times*, February 16, 2012, sec. Magazine. https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html.

Go Government. “Working in the Legislative Branch.” Go Government. Accessed September 24, 2022. https://gogovernment.org/all-about-government-jobs/working-in-the-legislative-branch/.

Godinho de Matos, Miguel. “Consumer Consent and Firm Targeting After GDPR: The Case of a Large Telecom Provider.” *Management Science* 68, no. 5 (2021): 3330–78.

Leswing, Kif. “Twitter Hackers Who Targeted Elon Musk and Others Received $121,000 in Bitcoin, Analysis Shows.” CNBC. Accessed September 24, 2022. https://www.cnbc.com/2020/07/16/twitter-hackers-made-121000-in-bitcoin-analysis-shows.html.

National Conference of State Legislatures. “State Laws Related to Digital Privacy.” Accessed December 10, 2022. https://www.ncsl.org/research/telecommunications-and-information-technology/state-laws-related-to-internet-privacy.aspx.

Padilla, Alex. “Statement of Vote: General Election November 3, 2020.” Secretary of State of California, November 3, 2020.

Pitofsky, Robert, Sheila F Anthony, Mozelle W Thompson, Orson Swindle, and Thomas B Leary. “Federal Trade Commission.” *PRIVACY ONLINE*, n.d., 208.

Rebecca Kern and Eric Geller. “Twitter Whistleblower to Congress: Your Data Is at Risk Too.” POLITICO. Accessed September 19, 2022. https://www.politico.com/news/2022/09/13/whistleblower-zatko-testimony-agrawal-twitter-00056291.

Sam Sabin. “States Are Moving on Privacy Bills. Over 4 in 5 Voters Want Congress to Prioritize Protection of Online Data.” *Morning Consult* (blog), April 27, 2021. https://morningconsult.com/2021/04/27/state-privacy-congress-priority-poll/.

Staszkiw, Michael, and Anna Mercado Clark. “Navigating Different Obligations of State Privacy Statutes for 2023.” *Rochester Business Journal* 38, no. 21 (October 21, 2022): 44.

The White House. “President Biden’s Bipartisan Infrastructure Law.” The White House. Accessed December 10, 2022. https://www.whitehouse.gov/bipartisan-infrastructure-law/.

Thorin Klosowski. “The State of Consumer Data Privacy Laws in the US (And Why It Matters).” Reviews for the Real World, September 6, 2021. https://www.nytimes.com/wirecutter/blog/state-of-privacy-laws-in-us/.

U.S. Department of Labor. “Guidance on the Protection of Personal Identifiable Information.” Accessed September 19, 2022. https://www.dol.gov/general/ppii.

Voss, W Gregory. “THE CCPA AND THE GDPR ARE NOT THE SAME: WHY YOU SHOULD UNDERSTAND BOTH.” *Competition Policy International*, 2021, 8.

Woodie, Alex. “Big Growth Forecasted for Big Data.” Datanami, January 12, 2022. https://www.datanami.com/2022/01/11/big-growth-forecasted-for-big-data/.

7. Appendix

7.1 Data Privacy Tolerance Score Feature Engineering

The data privacy tolerance score was calculated using answers to two umbrella questions. The first set of questions are listed below:

1. Thinking about your data, such as emails, photos, and other files that you put on the internet… How concerned are you that companies that provide these services would **keep a copy of your files even if you try to delete them**?
2. Thinking about your data, such as emails, photos, and other files that you put on the internet… How concerned are you that companies that provide these services would **use your photos in and other information in marketing campaigns?**
3. Thinking about your data, such as emails, photos, and other files that you put on the internet… How concerned are you that companies that provide these services would **not properly secure your personal information**?
4. Thinking about your data, such as emails, photos, and other files that you put on the internet… How concerned are you that companies that provide these services would **track your location using your cell phone**?

The second set of questions are also listed below:

1. How concerned are you that some of the information you share on social networking sites might be accessed by or shared with each of the following without your knowledge? **[Third parties, such as advertisers or businesses]**

2. How concerned are you that some of the information you share on social networking sites might be accessed by or shared with each of the following without your knowledge? **[The U.S. Government]**
3. How concerned are you that some of the information you share on social networking sites might be accessed by or shared with each of the following without your knowledge? **[Foreign governments or agents of foreign governments]**
4. How concerned are you that some of the information you share on social networking sites might be accessed by or shared with each of the following without your knowledge? **[Hackers]**
5. How concerned are you that some of the information you share on social networking sites might be accessed by or shared with each of the following without your knowledge? **[Political campaigns]**
6. How concerned are you that some of the information you share on social networking sites might be accessed by or shared with each of the following without your knowledge? **[Law enforcement agencies]**

The average responses to each group of questions was averaged, then averaged together – resulting in the data privacy tolerance score. N/A values were dropped as a part of this process.

7.1 R Code and Reproduction

This research used RStudio software to clean, transform, and analyze the dataset. To view the code, please visit the repository in GitHub at the following URL: https://github.com/cjmasamitsu/data-privacy/blob/main/capstone-code.R

In addition to the code, a CSV file including the dataset is publicly available in the repository for reproduction.

8. Curriculum Vita

Casey (Neubauer) Masamitsu was born in Ames, Iowa in 1990. In 2013, he completed a Bachelor of Arts in Advertising with a minor in Psychology at Iowa State University. Casey is also a licensed financial advisor and maintains his Project Management Professional (PMP) certificate. More recently, Casey co-founded a 501(c)(3) organization, Creating For Justice. The mission of Creating for Justice is to amplify BIPOC voices while fostering community, and facilitating accessible arts and professional education.

Casey currently works as a Senior Program Manager, Risk & Compliance at The New York Times. He will complete his Master of Science in Data Analytics & Policy at Johns Hopkins University in December 2022, after which he hopes to continue leveraging data analytics to uphold The New York Times' mission – to seek the truth and help people understand the world.