Essays in Household Finance and Consumer Protection

HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the Committee for the PhD in Business Economics have examined a dissertation entitled

Essays in Household Finance and Consumer Protection

Presented by **Talia B. Gillis**

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature

John Beshears, Chair

Signature

David Laibson

Signature

Brigitte Madrian

Signature

Gautam Rao

Date: May 5, 2022

Essays in Household Finance and Consumer Protection

A dissertation presented

by

Talia B. Gillis

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Business Economics

Harvard University

Cambridge, Massachusetts

May 2022

Dissertation Advisor:
Professor John Beshears

Author:
Talia B. Gillis

Essays in Household Finance and Consumer Protection

Abstract

This dissertation studies consumer decision-making and how household outcomes are shaped by technological change and regulatory context.

Chapter 1 (with John Beshears and Kartik Vira) approaches the study of household finance by analyzing household consumption decisions as they manage the day-to-day ebbs and flows of their liquid assets and debts. Using high-frequency account-level data from a retail bank in Australia, we show that consumption declines in the week after households make their monthly mortgage payment. Focusing on households that receive their income on a fortnightly basis, consumption drops more sharply over the course of a pay cycle when the cycle contains a mortgage payment than when the cycle does not contain a mortgage payment. This behavior is inconsistent with a neoclassical model augmented by incorporating liquidity constraints and is most consistent with models of present bias or mental accounting.

Chapter 2 (with Jens Frankenreiter and Dan Svirsky) focuses on consumer data usage by online service providers. We argue that privacy policies play an important role not only in describing current data practices but also in allocating among websites and consumers the power to decide whether a website can use consumer data in novel ways. By adapting standard models of incomplete contracts to privacy policies we explain the role of policies in allocating residual data rights and the potential role of regulation in limiting the allocation of rights to firms, similar to the approach of the E.U.'s General Data Protection Regulation (GDPR). We then use the model to explain how U.S. firms reacted to the GDPR, showing that U.S. websites with E.U. exposure are more likely to change their U.S. privacy policies

to have more permissive modification rules akin to allocating residual data rights to firms.

Chapter 3 (with Jann Spiess) discusses tensions between old law and new technology in the context of consumer credit markets. The ability to distinguish between people in setting the price of credit is often constrained by antidiscrimination laws. These laws were developed when human discretion played a central role in credit allocation and may be ineffective as pricing increasingly relies on intelligent algorithms that extract information from big data. Using a simulation exercise based on real-world mortgage data, we highlight the shortcomings of legal approaches that focus on algorithmic inputs and pricing rule interpretation. Instead, we argue that testing pricing outcomes is a promising path forward for evaluating discrimination claims in the algorithmic setting.

Contents

List of Tables

List of Figures

Acknowledgments

I am grateful to my dissertation committee for their guidance and support. David Laibson instilled in me a love for behavioral economics and continues to inspire me with his dedication to rigorous research and his masterful teaching. Brigitte Madrian provided incredible feedback on my work and invaluable life advice. Her enthusiasm for research and passion for improving public policy are contagious. Gautam Rao provided insightful guidance and support while encouraging me to not lose sight of the things that truly matter. I'm incredibly grateful to my committee chair, John Beshears, whose brilliance and sharpness are matched only by his patience and kindness. His clarity and depth of thought have guided all my work.

I also benefitted greatly from the advice and guidance of several faculty members in the Economics Department and Harvard Business School. I am especially grateful to Oliver Hart for his generosity in helping me with Chapter 2. I also want to express my gratitude to the friends who got me through graduate school, lifting me from the depths of Littauer's basement and grounding me from the 4th floor of Baker.

Finally, I am grateful to my family for their love and support. Most importantly, to my partner, Avi. We arrived in Cambridge just the two of us and left as a family of five.

Introduction

The financial lives of households are shaped by the interconnected relationship between consumer decision-making, firm behavior and regulatory structure. Recent research on how people actually make financial decisions and the dramatic technological changes in consumer markets make household finance a central domain in which to study the relationship between consumer, market and law. In this dissertation I attempt to consider all three perspectives. Focusing on household decision-making, I study how households manage their daily finances and the broader implications for modelling household consumption and savings behavior. From the firm's perspective, I study the way in which technological change, such as the use of advance statistical methods and big data, changes the ways in which firms make decisions in credit allocation and consumer data collection and usage. Finally, I consider how (and whether) law responds to changes in our understanding of consumer decision-making and technological advancement. Importantly, in the dynamic context of consumer financial markets, I also study the ways in which firms and consumers respond to the legal and regulatory framework.

In the first chapter (with John Beshears and Kartik Vira), we consider whether household consumption responds to the timing of a mortgage payment. A mortgage repayment is a frequently recurring outflow paid repeatedly, sometimes for decades. Most neoclassical theories of household finance would not predict a consumption response to this type of expected payment. Prior studies have documented consumption responses to predictable and recurring income increases; however, consumption responses to predictable and recurring income decreases have been understudied. Responses to anticipated outflows are particularly interesting because liquidity-constrained households can save in anticipation of such outflows and should smooth consumption at the time of the payment. Consumption responses to predictable and recurring outflows have important implications for understanding and modeling household financial decisions, which in turn can inform stimulus policies and policies that aim to increase household savings.

Using high-frequency account-level data from a retail bank in Australia, we show that consumption declines in the week after households make their monthly mortgage payment. Focusing on households that receive their income on a fortnightly basis, consumption drops more sharply over the course of a pay cycle when the cycle contains a mortgage payment than when the cycle does not contain a mortgage payment. The decrease in spending is concentrated in debit rather than credit card spending, even for households that pay interest on their credit cards. This behavior is inconsistent with a neoclassical model augmented by incorporating liquidity constraints and is most consistent with models of present bias or mental accounting.

In the second chapter (with Jens Frankenreiter and Dan Svirsky), we focus on privacy policies and how they allocate rights over data usage. As consumer data is increasingly used as a form of currency to access online services, it is important to understand how websites and consumers tradeoff the harms and benefits of data usage and the role of regulation.

While privacy policies' most important function is to provide consumers with a description of online service providers' current privacy practices, we argue that these policies also serve a second, often-overlooked function: They allocate among online services and consumers the power to decide whether a service can modify its privacy practices and use consumer data in novel ways. The allocation of residual data rights has been a central focus of the European Union's General Data Protection Regulation (GDPR), one of the most comprehensive and far-reaching privacy regulatory regimes, as it restricts policies from allocating broad rights for future data usage to service providers.

By adapting standard models of incomplete contracts to privacy policies to account for how consumers and firms allocate residual data rights and the use of broad data usage rights in privacy policy. We further develop a behavioral incomplete contracts model in which consumers discount future privacy harm. In this model, consumers may be worse off by agreeing to firms obtaining residual data rights than if they would retain these rights. This provides a theoretical explanation for this type of regulatory intervention adopted by the GDPR, which does not intervene with the substance of consumer data usage but only with the allocation of residual rights.

We then use the model to explain how U.S. firms reacted to the GDPR. We show that U.S. websites with E.U. exposure are more likely to change their U.S. privacy policies to have less stringent and more lenient modification rules. Among websites that do not have E.U. exposure, we see the opposite trend. These results suggest that websites sought to obtain residual rights over data usage after learning of the impact of the GDPR's stricter requirements.

In the third chapter (with Jann Spiess), we consider the implications of algorithmic consumer credit pricing. For many financial products, such as loans and insurance policies, companies distinguish between people based on their different risks and returns. However, the ability to distinguish between people by trying to predict future behavior or profitability of a contract is often restrained by legal rules that aim to prevent certain types of discrimination. These legal requirements were developed to challenge human discretion in setting prices and provide little guidance in a world in which firms set credit terms based on sophisticated statistical methods and a large number of factors.

In this chapter, we bring together existing legal requirements with the structure of machine learning decision-making in order to identify tensions between old law and new methods and lay the ground for legal solutions. We argue that while automated pricing rules provide increased transparency, their complexity also limits the application of existing law. Using a simulation exercise based on real-world mortgage data to illustrate our arguments, we note that restricting the characteristics that the algorithm is allowed to use can have a limited effect on disparity and can in fact increase pricing gaps. Furthermore, we argue that there are limits to interpreting the pricing rules set by machine learning that hinders the application of existing discrimination laws.

We end by discussing a framework for testing discrimination through a statistical analysis of the resulting algorithmic prices in a controlled environment. Unlike the human decision-making context, this framework allows for ex-ante testing of price rules, facilitating comparisons between lenders. We are that this type of analysis is especially useful for regulators that enforce antidiscrimination law.

Chapter 1

Consumption Responses to Mortgage Payments

Abstract

Using high-frequency account-level data from a retail bank in Australia, we show that consumption declines in the week after households make their monthly mortgage payment. Focusing on households that receive their income on a fortnightly basis, consumption drops more sharply over the course of a pay cycle when the cycle contains a mortgage payment than when the cycle does not contain a mortgage payment. The decrease in spending is concentrated in debit rather than credit card spending, even for households that pay interest on their credit cards. This behavior is inconsistent with a neoclassical model augmented by incorporating liquidity constraints and is most consistent with models of present bias or mental accounting.

1.1 Introduction

Does household consumption respond to the timing of a mortgage repayment? A mortgage repayment is a frequently recurring outflow paid repeatedly, sometimes for decades. Most neoclassical theories of household finance would not predict a consumption response to this type of expected payment. Prior studies have documented consumption responses to predictable and recurring income *increases*; however, consumption responses to predictable and recurring income *decreases* have been understudied. Responses to anticipated outflows are particularly interesting because liquidity-constrained households can save in anticipation of such outflows and should smooth consumption at the time of the payment. Consumption responses to predictable and recurring outflows have important implications for understanding and modeling household financial decisions, which in turn can inform stimulus policies and policies that aim to increase household savings.

In this paper we document a consumption decline after households pay their monthly mortgage. Focusing on spending on food, drink and entertainment away from home—spending that can be linked to instantaneous consumption—we show that a household's daily spending decreases in the week following a mortgage payment. This effect is concentrated in debit rather than credit card spending, even for households that actively use both debit and credit cards. Importantly, the consumption decline following a mortgage repayment is distinguishable from the consumption decline over a pay-cycle. Controlling for household and calendar date fixed effects, average household consumption declines over the course of a fortnightly (i.e., once-every-two-weeks) pay-cycle. However, in fortnights in which a mortgage payment is made, average daily consumption drops to a larger extent following the mortgage payment relative to fortnights without a mortgage payment. This finding provides evidence that the mortgage payment itself leads to a drop in consumption.

Our analysis relies on de-identified data from the Commonwealth Bank of Australia (CBA), one of Australia's largest retail banks. The dataset is based on the universe of CBA customers who have a mortgage with the bank and covers the years 2017-2018. Consumer mortgage accounts are linked to CBA savings and checking accounts, and the dataset includes debit and credit card transactions when consumers use a card issued by CBA. We restrict our sample to households who pay their mortgage on a monthly basis, receive their primary income deposited in a CBA bank account on a fortnightly basis, and use CBA debit and credit cards. We are left with a sample of 46,832 households.

In the first part of the paper, we study spending over the monthly mortgage payment cycle. Because we are primarily interested in consumption responses to mortgage payments, we focus on debit and credit card spending that is likely to reflect instantaneous consumption rather than spending on durable goods that could be consumed at a future date. Focusing on food, drink and entertainment away from home, we show that household average daily consumption drops by 0.06 Australian dollars (AUD) in the week following a mortgage payment. Relative to the median daily spending on food, drink and entertainment of 8.76 AUD, this decrease is a 0.7% drop. Consumption is highest in the week just before the mortgage payment.

We also find that the decline in consumption is concentrated in debit card spending and not credit card spending. Our main analysis focuses on consumers who formally have both a debit card and a credit card, even if they are not active users of both cards. We therefore also consider the subset of consumers who actively use both their debit card and their credit card, and we similarly observe a decline in consumption only on their debit cards. The decline in spending on debit rather than credit cards is particularly surprising as credit card spending tends to be more expensive than debit card spending.

In the second part of the paper, we focus on consumption over the fortnightly income cycle. Similar to previous studies (Mastrobuoni and Weinberg (2009); Olafsson and Pagel (2018)), we find that consumption declines over the pay-cycle. However, this decline is steeper for pay-cycles with a mortgage payment. Within a pay-cycle that includes a mortgage payment, we find that the difference in consumption before versus after the mortgage payment is greater than the difference in consumption between the equivalent sets of days in the previous pay-cycle (which does not include a mortgage payment).

We then discuss our findings in light of leading theories of household financial decision making. A useful benchmark is the lifecycle/permanent income hypothesis (LCH), according to which households make current consumption decisions based on their saving and consumption plan over their life cycle. A central prediction of the theory is that households aim to smooth consumption over their life cycle. Their consumption should not respond to predictable income changes because they can manage consumption path by saving and borrowing.[1]

Early studies of the LCH documented the ways in which consumers do in fact respond to predictable changes in income.[2] These findings inspired a modification of the original LCH theory, known as the buffer stock model, that incorporated liquidity constraints that prevent consumers from smoothing consumption across periods.[3] Liquidity constraints limit the extent to which individuals are able to borrow against future income, so consumers may under-consume during periods of lower income and increase consumption with income. Empirical studies have considered one-time infusions of income, such as tax refunds,[4] other types of transfers[5] or an adjustment of income.[6] These studies document a spending response at the time of the cash inflow rather than the time at which consumers receive news of the expected income. While not all results are easily explained by the existence of liquidity constraints,[7] they support the notion of frictions in consumption smoothing.

Our findings cannot be explained by the buffer-stock model. Liquidity constraints refer

[1]See survey article Gomes *et al.* (2021)

[2]Many of these studies focused on one-off events of changes in income. Although these changes were one-off or rare, they were still predictable, suggesting that any spending response should occur when learning of these changes and not at the actual time of money receipt. See, e.g., Hall (1978); Zeldes (1989); Johnson *et al.* (2006)

[3]See, e.g., Carroll (1997); Kaplan and Violante (2014)

[4]See, e.g., Souleles (1999); Shapiro and Slemrod (2003)

[5]See, e.g., Hsieh (2003) and Kueng (2018), which study the Alaska Permanent Fund. See also Johnson *et al.* (2006) and Parker *et al.* (2013) study the consumption responses to the government stimulus program in 2001 and 2008, respectively.

[6]For example, a predictable change in income as a result of a tax policy change. Parker (1999) documents a consumption response to a change in Social Security tax withholding.

[7]Various studies have focused on liquidity constrains as an explanation for the high propensity to spend out of an inflow. See, e.g., Bodkin (1959); Zeldes (1989); Johnson *et al.* (2006).

to limitations on borrowing against future income, but households can save in anticipation of a mortgage payment to smooth their consumption.[8] The consumption response we document to an income *decline* suggests that liquidity constraints cannot fully explain violations of the consumption smoothing prediction.

Recent studies that consider income cycles, rather than one-time infusions of income, have highlighted limitations of the buffer-stock model and have emphasized behavioral explanations for household consumption. These studies use higher frequency and more detailed spending data than earlier studies. The main finding of this literature is that even households with meaningful cash-on-hand respond to income receipt (Olafsson and Pagel (2018)).[9] Importantly, some of these studies show a repeated response to predictable and recurring income, such as a consumption response to receiving a regular paycheck.[10] These studies, as well as studies that show a consumption decline over the income cycle (Mastrobuoni and Weinberg (2009); Shapiro (2005); Hastings and Shapiro (2018)), have motivated more behavioral theories of household consumption, suggesting that household impatience is driving consumption patterns.

We consider our findings in light of a model of hyperbolic discounting (Laibson (1997); Angeletos *et al.* (2001)). Following the analysis in Harris and Laibson (2001), who derive a Hyperbolic Euler Relation, we consider a model in which a household's marginal propensity to consume is sensitive to household cash-on-hand, which is lower in income pay-cycles with a mortgage payment. Households that expect to have lower liquid assets in the future also expect that they will overconsume later in the income cycle.[11] Therefore, households

[8]Shea (1995) discuss how liquidity constraints would lead to more sensitivity to income increases than to income declines. Baugh *et al.* (2021) document an asymmetric response to tax refunds and tax payments and discuss how liquidity constraints would not predict a response to a tax payment, similar to their empirical results.

[9]Gelman *et al.* (2014) study increases in spending at the time of income receipt and argue that it is explain by regular payments and not consumption.

[10]See, e.g., McDowall (2019) documenting a short-lived spending response to the receipt of a paycheck.

[11]Assuming that households are sophisticated about predicting future behavior. See Harris and Laibson (2001). For a discussion on the difference between being a sophisticated and naïve agent with respect to beliefs about future discounting, see O'Donoghue and Rabin (1999)

may choose to consume more earlier in the cycle. When households have more liquid assets, such as in pay-cycles without a mortgage payment, they predict a lower marginal propensity to consume later in the pay-cycle, leading to lower levels of consumption early in the cycle. This creates a steeper consumption decline over cycles with mortgage payments, consistent with our results.

Mental accounting models may also explain our findings. In models of mental accounting, households use rules-of-thumb to determine their consumption paths. In these models, the fungibility of interchangeable resources breaks down because individuals create different mental accounts. In our setting, it is possible that consumption decisions are made based on a rule-of-thumb that treats current spendable income, reflected by a household's bank account balance, as separate from other funds. When the bank account balance is lower than normal, households decrease consumption, even if the lower balance reflects a predictable change in the ebbs and flows of the account. Our finding of a response in debit card spending, but not in credit card spending, to mortgage payments supports a mental accounting explanation for consumption responses. If mortgage payments impact their bank account balances, households may choose to adjust spending on debit cards, which are directly linked to bank account balances.

Our paper contributes to the literature documenting consumption responses to income *declines*, which has mixed findings. Gelman *et al.* (2020) study the 2013 U.S. Federal government shutdown and show that employees were able to smooth consumption during the temporary 2-week drop in income. Similarly, Baugh *et al.* (2021) demonstrate that households smooth consumption when making a tax payment. Other studies, however, have documented consumption decreases in response to predictable income declines. Ganong and Noel (2019) study the exhaustion of unemployment benefits, a predictable change in income stream, and document a decline in consumption after the end of unemployment benefits.

The study of income declines has so far focused on limited settings in which income declines tend to be one-off or rare events. While a tax payment is likely to be more frequent than the exhaustion of unemployment benefits, it is still infrequent (no more than once a year), and the exact amount of the payment varies. Even if in theory these payments are predictable, some consumer inattention is likely to lead to a consumption response to the income decline. An income change at the end of unemployment benefits may be more salient and predictable than a tax prepayment, but it is still a rare event for most workers, potentially coinciding with many other life circumstances that shape behavior. Unlike consumption responses to cash *inflows* that arrive in a recurring cycle (e.g., income that arrives in a recurring pay-cycle), consumption responses to cash *outflows* that occur on a regular schedule have yet to be studied.

Our context provides an ideal setting in which to study consumption responses to recurring cash outflows. We study the decline in available resources that is a result of a mortgage payment, which tends to be roughly stable in size across time. For most of the mortgages in our sample, monthly amounts remain similar sometimes for decades, adjusting income in a consistent way. This creates a highly salient and recurring payment, which households experience time and time again, so a consumption response to this payment suggests household behavior unaccounted for in many standard models.

The paper proceeds as follows. In Section 1.2 we provide institutional details of the context and describe our data. We also discuss our sample restrictions and definitions of outcome variables. Section 1.3 shows that consumption is sensitive to the timing of a mortgage payment, declining in the week following a payment. Section 1.3 also demonstrates that within a pay-cycle there is a drop in consumption after a mortgage payment and that this drop is greater than the equivalent drop in pay-cycles without a mortgage payment. Section 1.4 evaluates these findings in light of theories commonly used to explain household consumption.

1.2 Data

1.2.1 Institutional Setting and Data

The data we use comes from the Commonwealth Bank in Australia (CBA), one of the largest retail banks in Australia. Retail banking is fairly consolidated in Australia with over the largest four banks serving over 70% of consumers. Banking is Australia is considered centralized because many costumers use their retail bank for financial services beyond deposits, such loans (including mortgages), payments cards, investment services and insurance. Given the concentration of the Australian banking industry, consumers are more likely receive a mortgage from their retail bank and are more likely to use debit and credit cards issued by their retail bank than is typical in the U.S. Overall, banks in Australia are the primary mortgage originators, with largest four retail banks holding around 80% of mortgages.

Our data consists of three types of datasets. First, we have data on costumer mortgage accounts at CBA. This includes information at the time of mortgage origination, such as the date of funding, loan amount and loan term. We also have information on the mortgage over time such as the mortgage balance, repayment dates and repayment amounts. Our second dataset covers information on costumer bank accounts. This dataset covers account balances and account transactions. Importantly, we use this dataset to infer information about household income. Lastly, we use a dataset that covers information on debit card and credit card spending, including information on balances and repayments. We observe daily transaction amounts for each merchant category, which are spending categories set by CBA based on vendor type at which a payment card was used.

1.2.2 Analysis Sample

Our analysis sample is drawn from households with mortgages at CBA from January 2, 2017 through December 31, 2018, which includes over 600,000 households. The unit of observation is household-by-day. We define households as the group of individuals who share an account at CBA and limit our sample to households of one or two individuals.

Our primary analysis imposes three main restrictions on our sample. First, we restrict our sample to households that receive their income on a fortnightly basis, meaning every 14 days, as defined below. In Australia, employers typically select the paycheck frequency, with a fortnightly income cycle being most common. Second, we restrict our sample to households that repay their mortgages at a monthly frequency.[12] Not all consumers pay their mortgage on a monthly repayment schedule. CBA allows households to opt-out of the default monthly repayment schedule, and select a weekly or fortnightly schedule, although most customers pay their mortgage on a monthly schedule. By focusing on consumers that repay their mortgage monthly, we can study within household effects of the alignment and misalignment of mortgages repayments and income.

Our third restriction is to limit the sample to households that use CBA as their primary bank account. We restrict our primary analysis to households for which we have data of income deposited in a CBA bank account and have an average of over 3 card transactions a week within the first three months of being included in our sample.[13] We are left with a sample of 46,832 households who repay their mortgage on a monthly schedule, receive their income on a fortnightly basis, and for who CBA is their primary financial institution.[14]

Some of our analysis distinguishes between debit card and credit card spending. While our full sample includes households who actively use either of their cards, for some of our analysis we wish to consider households who actively use both their debit and credit cards. For this subsample, we identify households that have at least one transaction on both their debit and card cards. This restriction creates a subsample 22,905 households who make transactions using both debit and credit cards, which we refer to as "debit and credit card active users."

[12]We define the repayment frequency according to the frequency of our first recorded repayment. If a household has multiple mortgages, we only look at the mortgage(s) that are funded on the household's earliest funding date.

[13]This would be either within the first three months for which we have data on the household, or within three months after mortgage funding, whichever is later.

[14]We drop observations from months where we do not have income data.

1.2.3 Variable Construction

Spending.—Spending is measured from debit and credit card transactions, within spending categories defined by CBA based on the vendor at which a payment card is used. We define "total amount spent" as the aggregate household spending across cards, a day.[15] We create a sub-group of spending, "food, drink, and entertainment away from home" representing non-durable and instantaneous consumption, which includes a limited number of spending categories which we consider purchases that are immediately consumed. This sub-group of spending allows us to attach the consumption experience to the observed time of card purchase, so that we are able to measure changes in consumption at a high frequency.

We refer to this category of spending as "consumption" even though it only captures a small portion of overall household consumption. With some vendor categories the goods or services purchased at the time takes the transaction takes place and recorded is not necessarily the time they are consumed. For example, for the category "hardware stores" durable goods might be consumed over months or years from the time of the transaction. With some vendors, a transaction may include goods that are instantaneously consumed such as with the vendor category "grocery stores." However, because we only observe the vendor and not the actual goods purchased, we are unable to distinguish between goods consumed immediately and goods consumed in the future. We therefore isolate vendors for which all purchases are likely to reflect immediate consumption and refer to this spending as "consumption." This sub-group includes spending in the categories of restaurants, bars, cafes and places of entertainment, such as bowling alleys and movie theaters.

Income cycle.—We look at all fortnights during which a household appears in the data set, and divide the data for a household into fortnightly periods. Within each household-by-fortnight period, we identify the two-day window with the highest fraction of income payments. Looking across fortnightly periods for a household, a household is defined to be receiving income on a *fortnightly basis* if at least 75% of fortnights have the same two-day

[15] We define spending as 0 on day in which no card transaction is recorded. We also winsorize the amount spent at the 1st and 99th percentiles.

window containing at least 75% of income for the fortnight.[16]

We calculate every calendar day's position in the income cycle for each household, relative to the payment window for the household. Each day receives a number from 0-12, with days in the two-day payment window being labelled 0.

Mortgage repayment cycle.—We calculate every calendar day's position in the mortgage repayment cycle for each household, relative to the nearest payment date, which we consider day 0. All other days in the repayment cycle receive a number from -15 to 15. Most of our analysis looks at bins of mortgage repayment cycle days rather than each day in the cycle separately. For example, in Section III we use bins of 7 days for our regression analysis and bins of 3 days for our graph.

Table 1.1 summaries our main outcome variables and sample for the analysis in the following sections. For our outcome variable of consumption—food, drink and entertainment away from home—the mean of daily spending is 11.70 Australian dollars (AUD) and the median daily spending is 8.76 AUD, which is roughly 8.86 U.S. dollars and 6.63 U.S. dollars, respectively, in 2017-2018.[17] For our subset of households that are debit and credit card active users the mean of average daily spending is 12.91 AUD and the median average daily spending is 10.06 AUD.

1.3 Consumption Responses to Mortgage Repayment

In this section, we show that consumption is sensitive to the timing of a mortgage payment with consumption declining in the week after households make their monthly mortgage payment. We also provide evidence that consumption drops more sharply over the course of a pay cycle when the cycle contains a mortgage payment than when the cycle does not contain a mortgage payment.

---

[16]Households with less than three fortnight observations are dropped from the sample.

[17]This is based on the average exchange rate of 1 AUD = 0.7669 USD in 2017 and an average exchange rate of 1 AUD = 0.7475 USD in 2018.

Table 1.1: *Summary Statistics*

	Mean	Std. dev.	Percentiles 1st	10th	50th	90th	99th
Original mortgage amount	399,966	240,103	42,542	150,000	357,013	695,149	1,348,776
Mortgage balance when first entering sample	327,227	233,731	23	47,609	300,000	615,001	1,199,560
Income from primary fortnightly source (annualized)	63,499	35,227	6,551	26,586	57,471	105,993	205,396
Total income (annualized)	68,097	38,242	7,122	28,397	61,207	114,532	221,703
Total bank account balances	43,758	84,463	-168	1,284	11,056	121,201	508,190
Mortgage principal repayment amount (annualized)	14,460	12,929	0	0	12,855	31,232	58,055
Daily spending on food, drink, and entertainment	11.70	10.54	0.06	1.84	8.76	25.25	56.11
Sample with debit and credit	12.91	10.93	0.30	2.45	10.06	26.98	57.89
Debit card spending only	5.04	6.65	0	0	2.45	13.71	33.76
Credit card spending only	7.74	9.22	0	0.07	4.42	20.11	45.74

This table reports summary statistics for the sample of 46,832 households used in the main analysis. There are 22,905 households that have both debit and credit cards. For each household and for each variable except the first two variables (original mortgage amount and mortgage balance when first entering sample), we calculate the within-household mean of the variable across all of the days on which the household is included in the sample, and we then calculate the summary statistics for these within-household means, weighting households equally. For all variables, we winsorize at the 1st and 99th percentiles before calculating the means and standard deviations reported in the table. All amounts are in Australian dollars.

1.3.1 Daily Consumption Drop after Mortgage Payment

We start by analyzing consumption over the monthly mortgage repayment cycle.

Methodology.—For each household, we calculate daily consumption, meaning spending on food, drink and entertainment away from home. To estimate the change in consumption over the monthly mortgage cycle we create four bins based on the distance from the mortgage repayment day. We define the mortgage repayment day as day 0, and create two bins covering the weeks *before* the mortgage repayments (days -15 through -8 and days -7 through -1) and two bins covering the week of mortgage repayment and *after* the mortgage repayment (days 0 through +6 and days +6 through +15). For each bin we calculate the average daily consumption.

Our analysis includes two samples. Our main sample considers all households that have a debit card and credit cards and our subsample includes households that are debit and credit card active users. For each of the two samples we consider their combined debit and credit card consumption, and debit card and credit card consumption separately. We use the following regression to estimate the impact of mortgage repayment on consumption:

$$consumption_{h,t,i,m} = \alpha_h + \tau_t + \eta_i + \theta_m + \epsilon_{h,t,i,m} \tag{1.1}$$

Where α_h is our household fixed effect, τ_t is our calendar day effect (t = 1, 2, . . . , 730), η_i is our income cycle day effect (i = 0, 1, . . . , 12) and θ_m is our mortgage repayment cycle effect (m in bins of roughly 7 days). Our household fixed effect (α_h) is meant to absorb average household-specific consumption. Our calendar day fixed effect (τ_t) controls for the average spending on a particular calendar day within our two-year period and is meant to absorb consumption fluctuations for weekends, holidays and seasons that could bias our results. Because consumption levels can change within a fortnightly income cycle, we control for these fluctuations using the income cycle day fixed effect (η_i). Our main independent variable, θ_m, is a set of dummy variables for each of the mortgage repayment cycle weeks, where the week immediately before the mortgage repayment (days -7 to -1) is the omitted variable.

Key Results.—Table 1.2 demonstrates that household consumption drops in the two weeks following a mortgage repayment with the largest drop in the first week after a mortgage payment. In our full sample, spending on food, drink and entertainment declines by 0.06 AUD in the week just after the mortgage repayment relative to the week before the mortgage repayment. Relative to the median daily average consumption of 8.76 AUD (see Table 1.1) this decrease is a 0.7% drop in the week after a mortgage repayment relative to the week before the mortgage repayment.

When restricting our analysis to households of debit and credit card active users, we see similar results with an average decline of 0.08 AUD in the week following the mortgage repayment. Relative to median daily average spending on food, drink and entertainment for

Table 1.2: *Daily Spending on Food, Drink, and Entertainment over the Mortgage Repayment Cycle*

Sample	Full Sample			Households that have both debit and credit cards		
Cards included in spending measure	Debit and credit	Debit cards only	Credit cards only	Debit and credit	Debit cards only	Credit cards only
Days in mortgage repayment cycle						
-15 through -8	-0.015	-0.011	-0.003	-0.048*	-0.029**	-0.014
	(0.015)	(0.007)	(0.012)	(0.021)	(0.011)	(0.016)
-7 through -1	--	--	--	--	--	--
0 through +6	-0.062***	-0.047***	-0.008	-0.083***	-0.070***	-0.002
	(0.015)	(0.006)	(0.013)	(0.022)	(0.011)	(0.016)
+7 through +15	-0.041***	-0.026***	-0.012	-0.049*	-0.042***	0.000
	(0.015)	(0.007)	(0.012)	(0.022)	(0.011)	(0.017)
Number of households	46,832	46,832	46,832	22,905	22,905	22,905
Number of observations	23,287,476	23,287,476	23,287,476	11,687,369	11,687,369	11,687,369

+ Significant at 10% level. * Significant at 5% level. ** Significant at 1% level. *** Significant at 0.1% level.

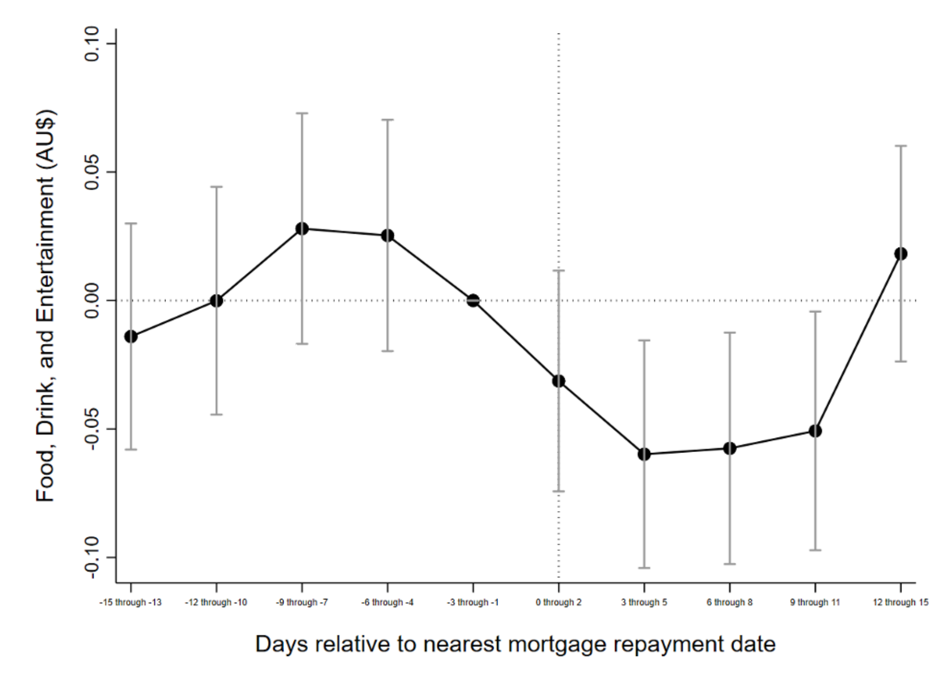
In this table, each column reports the results from an ordinary least squares (OLS) regression with the outcome variable and the sample named in the column header. Each outcome variable is defined at the daily level in Australian dollars and is winsorized at the 1st and 99th percentiles. The reported coefficient estimates and standard errors are for an indicator for days -15 through -8, an indicator for days 0 through +6, and an indicator for days +7 through +15 of the household's mortgage repayment cycle, where negative numbers denote the number of days before the nearest mortgage repayment date and positive numbers denote the number of days after the nearest mortgage repayment date (days -7 through -1 constitute the omitted category). The other explanatory variables are indicators for each day of the household's fortnightly income cycle (grouping together the two days identified as the window for receiving fortnightly income), indicators for each calendar date in the sample period, and household fixed effects. Standard errors with two-way clustering by household and by calendar date are in parentheses.

this subsample of 10.06 AUD (see Table 1.1) this decrease is a 0.8% drop relative to the week before the mortgage repayment.

The drop in consumption is concentrated in debit card spending and not credit card spending. Comparing debit card spending and credit card spending in our full sample reveals that consumption declines for debit card spending but not for credit card spending. We see a similar pattern in our subsample of households that debit and credit card active users. Although households in this sample are actively user both cards, the change in consumption is reflected primarily in debit card spending.

We also study narrower bins of 3 days to examine the change in consumption following a mortgage payment. Figure 1.1 shows the change in consumption over the mortgage repayment cycle, where day 0 is the mortgage repayment day. Similar to our regression results, consumption declines in the week following the mortgage payment and increases to its highest level in the week before a mortgage payment.

Figure 1.1: *Daily Spending on Food, Drink, and Entertainment over the Mortgage Repayment Cycle*



This figure shows coefficient estimates and the associated 95% confidence intervals from an ordinary least squares (OLS) regression with spending on non-durable, immediate consumption as the outcome variable. This variable is defined at the daily level in Australian dollars and is winsorized at the 1st and 99th percentiles. The sample is the 46,832 households used in the main analysis. The reported coefficient estimates and 95% confidence intervals are for an indicator for days -15 through -13, an indicator for days -12 through -10, an indicator for days -9 through -7, an indicator for days -6 through -4, an indicator for days 0 through +2, an indicator for days +3 through +5, an indicator for days +6 through +8, an indicator for days +9 through +11, and an indicator for days +12 through +15 of the household's mortgage repayment cycle, where negative numbers denote the number of days before the nearest mortgage repayment date and positive numbers denote the number of days after the nearest mortgage repayment date (days -3 through -1 constitute the omitted category). The other explanatory variables are indicators for each day of the household's fortnightly income cycle (grouping together the two days identified as the window for receiving fortnightly income), indicators for each calendar date in the sample period, and household fixed effects. The 95% confidence intervals are calculated from standard errors with two-way clustering by household and by calendar date.

The economic magnitude of the consumption decline in the week after a mortgage payment is not very large but it is surprising that the timing of an expected and recurring payment has an effect on consumption at all. As discussed in more detail in Section 1.4, neoclassic theories of household finance would not predict a consumption response at the time of a mortgage payment, even for households that are liquidity constrained. In contrast to the predictions of these theories, we find that even households with higher income drop their consumption in the week following a mortgage payment.[18]

The decrease in spending on debit rather than credit card spending is particularly surprising. If a household is paying interest on their credit card, households should cut back on their more expensive credit card spending. Instead, we find that households could be choosing to cut back on their less expensive debit card spending. Because some households treat their credit card as a transaction card and do not pay interest on their credit card spending, Appendix A.5 repeats our main analysis for a subsample of households that pay interest on their credit card (20,301 households). Even these households, which are paying more for their spending on credit cards than debit cards, the decline in consumption concentrated in their less expensive debit card spending.

Heterogeneity Analysis and Robustness.—The decline in consumption following a mortgage payment is robust to other specifications of our spending outcome variable. The Appendix reports results similar to Table 1.2 for a spending category that only includes food and drink away from home. The Appendix also reports the results of our analysis broken down by income and principal payment amount terciles as well as terciles of the ratio of income and principal amount.

1.3.2 Consumption Decline in Pay-cycles with Mortgage Payments

We now turn to study consumption over the fortnightly pay-cycle. We compare the consumption decline that takes place over a fortnightly pay-cycle with a mortgage payment to the consumption decline in pay-cycles without a mortgage payment. Prior literature has demonstrated that consumption can drop within a pay-cycle,[19] and so in this section we study whether the decline in consumption is different in weeks with a mortgage payment.

Methodology.—We start by calculating the average daily consumption before and after a mortgage payment within each pay-cycle that includes a mortgage repayment. We refer to

---

[18]See Appendix A.2 for a breakdown of our analysis by income tercile.

[19]See, e.g., Shapiro (2005), Hastings and Shapiro (2018) and McDowall (2019).

the difference in consumption before versus after the mortgage payment as the consumption "drop."

For each pay-cycle that includes a mortgage payment we select the preceding pay-cycle as the matched pay-cycle. Although this matched pay-cycle does not include a mortgage payment we calculate the average daily consumption before versus after the equivalent day of the cycle on which the mortgage is paid during the following pay-cycle. We think of this as the "pseudo" repayment day as it is the day that corresponds to the repayment day in the matched fortnight that includes a mortgage payment. This allows us to calculate the equivalent consumption drop for the matched pay-cycles which do not include a mortgage repayment.

Using our matched sample of pay-cycles, one of which includes a mortgage payment and one of which does not include a mortgage payment, we compare the consumption drop using a regression with an indicator for pay-cycles with a mortgage repayment:

$$consumption_drop_{h,p,c} = \alpha_{h,p} + \beta PaycycleWithMortgage_{h,p,c} + \epsilon_{h,p,c} \tag{1.2}$$

We indicate each pay-cycle by c and each matched pair of pay-cycles (two fortnights) by p. We use $\alpha_{h,p}$ as the household by matched pay-cycle fixed effect. Our main independent variable $PaycycleWithMortgage_{h,p,c}$ is an indicator variable equal to 1 if the pay-cycle includes a mortgage payment and 0 if the pay-cycle does not include a mortgage payment. Our coefficient β reflects the difference in consumption drop for pay-cycles that include a mortgage payment and pay-cycles that do not include a mortgage payment.

In our second analysis, we compare the average daily consumption before and after a mortgage repayment, in pay-cycles that include and do not include a mortgage repayment. We estimate the average daily consumption for four categories: 1. Consumption *before* a mortgage payment in a cycle <u>with</u> a mortgage payment; 2. Consumption *after* a mortgage payment in a cycle <u>with</u> a mortgage payment; 3. Consumption *before* a pseudo mortgage payment in a matched cycle <u>without</u> a mortgage payment; 4. Consumption *after* a pseudo mortgage payment in a matched cycle <u>without</u> a mortgage repayment.

To study the average daily consumption for these four categories we estimate the following regression:

$$\begin{aligned} consumption_{h,t,c} = \alpha_h + \tau_t + \beta_1 PaycycleWithMortgage_{h,c} + \beta_2 AfterRepayment_{h,c,t} \\ + \beta_3 \left(PaycycleWithMortgage_{h,c} * AfterRepayment_{h,c,t} \right) + \epsilon_{h,t,c} \end{aligned} \tag{1.3}$$

Where α_h is our household fixed effect and τ_t is our calendar day effect (t = 1, 2, . . . , 730). The dependent variable $PaycycleWithMortgage_{h,c}$ is an indicator equal to 1 if the day of consumption is in a pay-cycle that includes a mortgage payment and equal to 0 if the day of consumption is in a pay-cycle without a mortgage payment. The dependent variable $AfterRepayment_{h,c,t}$ is an indicator for whether the day of consumption is before versus after a mortgage payment (in pay-cycles that include a mortgage payment) or a pseudo mortgage payment (in pay-cycles that do not include a mortgage payment). This indicator variable is equal to 1 for days after a mortgage payment (in pay-cycles that include a mortgage payment) or a pseudo mortgage payment (in pay-cycles that do not include a mortgage payment), and equal to 0 if the day of consumption is before a mortgage payment or pseudo mortgage payment. The regression coefficients allow us to estimate the average daily consumption in our four categories.

Key Results.—Table 1.3 shows that the decline in consumption over the pay-cycle is greater in pay-cycles that include a mortgage payment than pay-cycles that do not include a mortgage payment. Using regression 1.2 we estimate that the drop in average daily consumption in pay-cycles that include a mortgage payment is 0.06 AUD greater than the equivalent drop in pay-cycles that do not include a mortgage payment. Similar to our findings in Section 1.3.1, the larger drop in consumption for pay-cycles that include a mortgage repayment is concentrated in debit card spending and not credit card spending.

These results are consistent with a consumption response to the mortgage payment itself. As we discuss in the next section, this greater drop in consumption could be explained by a model of hyperbolic discounting in which the effective discount rate is sensitive to a household's liquid assets. In pay-cycles that include a mortgage payment, households

Table 1.3: *Difference in Daily Spending on Food, Drink and Entertainment Away from Home Between Income Cycles with and without Mortgage Payment*

Sample	Full Sample			Households that have both debit and credit cards		
Cards included in spending measure	Debit and credit	Debit cards only	Credit cards only	Debit and credit	Debit cards only	Credit cards only
1(fortnights with a mortgage payment)	-0.061*	-0.051***	-0.002	-0.042	-0.060**	0.025
	(0.027)	(0.012)	(0.022)	(0.039)	(0.018)	(0.031)
Number of observations	1,281,768	1,281,768	1,281,768	643,806	643,806	643,806

+ Significant at 10% level. * Significant at 5% level. ** Significant at 1% level. *** Significant at 0.1% level.

have lower cash-on-hand which may lead to higher consumption earlier on in the pay-cycle resulting in a steeper decline in consumption over the pay-cycle.

Table 1.4 summarizes the average daily consumption for the four categories of consumption before and after a mortgage payment or pseudo mortgage payment, in pay-cycles that include a mortgage payment and pay-cycles that do not include a mortgage payment, using the estimates of regression 1.3.

Table 1.4: *Daily Spending on Food, Drink and Entertainment Away from Home in Income Cycles with and without Mortgage Payment and before and after Mortgage Payment*

Sample	Full Sample			Households that have both debit and credit cards		
Cards included in spending measure	Debit and credit	Debit cards only	Credit cards only	Debit and credit	Debit cards only	Credit cards only
Before payment, w/o mortgage	--	--	--	--	--	--
Before payment, w/ mortgage	0.014	0.011	0.002	0.001	0.012	-0.013
	(0.017)	(0.007)	(0.013)	(0.024)	(0.013)	(0.018)
After payment, w/o mortgage	-0.087***	-0.100***	0.029+	-0.120***	-0.137***	0.030
	(0.018)	(0.009)	(0.015)	(0.027)	(0.014)	(0.021)
After payment, w/ mortgage	-0.155***	-0.150***	0.019	-0.193***	-0.201***	0.031
	(0.018)	(0.009)	(0.015)	(0.027)	(0.014)	(0.021)
Number of households	46,500	46,500	46,500	22,829	22,829	22,829
Number of observations	21,010,106	21,010,106	21,010,106	10,553,418	10,553,418	10,553,418

+ Significant at 10% level. * Significant at 5% level. ** Significant at 1% level. *** Significant at 0.1% level.

The average daily consumption after a mortgage payment is lower by 0.17 AUD than in the days before the mortgage payment. In pay-cycles that do not include a mortgage repayment, the average daily consumption after the pseudo payment day lower by 0.09 AUD than the days before the pseudo repayment day.

Together, these results demonstrate that the timing of a mortgage payment impacts the average daily consumption and that this impact is distinct from the general consumption decline over the pay-cycle. In the next section we discuss a number of leading households finance theories and whether they are consistent with our findings.

Heterogeneity Analysis and Robustness.—When using a narrower outcome variable of spending on food and drink away from home, we find a similar larger drop for pay-cycles with a mortgage payment (see Appendix). The Appendix also reports our results broken down by bank account tercile.

1.4 Discussion and interpretation

This section considers our results in light of the predictions of main theories of household behavior that have been used to explain previously observed responses to predictable income changes.

1.4.1 Buffer-Stock Model

In a buffer-stock model, households select their consumption path based on their lifetime budget constraint, which would take into account uncertain future income and liquidity constraints. In models of household consumption with liquidity constraints, households may respond to expected increases in income. While households may aim to smooth consumption temporally, if households are constrained in their ability borrow in anticipation of future income, they may be consuming below their optimal unconstrained consumption path.[20]

[20]See, e.g., Zeldes (1989) which includes borrowing constraints in a model of infinite horizon consumption allocation.

This could lead to increased consumption at the time of income receipt when those funds become available.

Even if households cannot borrow against future income, they have the ability to save in preparation for an anticipated cash outflow. The buffer-stock model therefore would predict no consumption decline following a mortgage payment.[21] In contrast with the buffer-stock model's prediction we observe a decline in consumption following a predictable and recurring cash outflow.

1.4.2 Present Bias

Models with behavioral households typically assume that households are more impatient than standard exponential discount rates or are not fully forward looking. A well-known example of a behavioral household is a hyperbolic discounting household (Laibson (1997)) which is present bias.

A highly present bias household may choose a higher consumption path at times closer to income receipt and then decrease consumption over time, in a manner which is close to what Campbell and Mankiw (1989) describe as "hand-to-mouth" behavior.[22]

In classic formulations of hyperbolic discounting[23] households select their consumption path based on their lifetime budget constraint, similar to a household in a buffer-stock model, which reflects the net income of the household including both known cash inflows and known cash outflows. Assuming that a mortgage payment is expected, the model would predict that the consumption path is not affected by the exact timing of the mortgage repayment.

A household that allocates consumption across a fortnightly income cycle would need

[21] See, e.g., Shea (1995) shows that the existence of liquidity constraints would predict more of a response to income increases than decreases. However, Shea (1995) documents a greater response to predictable income declines.

[22] For an analysis of wealthy hand-to-mouth households, see Kaplan and Violante (2014).

[23] See, for e.g. Angeletos *et al.* (2001) in which households have both a short-run preference for instantaneous gratification and a long-run preference to act patiently. This is why the model is known as "hyperbolic discounting."

to account for the mortgage payment in its budget constraint. If a household faces liquidity constraints or income uncertainty, whether an income cycle includes a mortgage payment can impact the budget constraint for consumption within an income cycle. However, the timing of the mortgage should not impact the consumption path within the income cycle. In contrast, our results suggest consumption sensitivity to the timing of a mortgage payment within a pay-cycle.

While some models of hyperbolic discounting would not predict that the timing of a mortgage payment within an income cycle should impact the consumption path in that income cycle, we are unable to fully distinguish between income cycle effects and mortgage payment effects. Because mortgage payments are monthly and income is received fortnightly, the decline in consumption in the week after a mortgage payment, discussed in 1.3.1, may also be capturing the pay-cycle consumption decline.[24] In Section 1.3.2 we attempt to more directly separate between the effect of mortgage payment and the pay-cycle consumption decline and show that the decline in consumption is indeed greater for pay-cycles with a mortgage payment. This demonstrates that the decline in consumption in the week following a repayment is not explained by pay-cycle consumption decline.

The greater decline in consumption in pay-cycles with a mortgage repayment may, however, be explained by a hyperbolic discounting model if a shift in the budget constraint leads to a steeper decline in consumption. Harris and Laibson (2001) derive a Hyperbolic Euler Relation in which the effective discount rate varies depending on future cash-on-hand. Because lower future liquid wealth increases the marginal propensity to consume (MPC) in the future, households will consume more in the present in order to prevent overconsumption by their future self.[25] Effectively the higher future MPC leads to a discount rate closer to the short run discount rate ($\beta\delta$) rate than an exponential discount rate (δ).

---

[24] As discussed above, this decline over an income cycle has been document by Shapiro (2005), Hastings and Shapiro (2018) and Mastrobuoni and Weinberg (2009).

[25] This derivation assumes that agents are sophisticated with respect to their futures selves. For further discussion, see O'Donoghue and Rabin (1999) who distinguish between naïve agents that believe that will behave in a time-consistent manner in the future and sophisticated agents that are aware that in the future they will also be present bias.

In pay-cycles that include a mortgage payment households have lower liquid wealth which could lead to a higher MPC.[26] This higher future MPC will increase consumption earlier in the pay-cycle as the effective discount rate is closer to the short run discount rate $\beta\delta$. In pay-cycles that do not include a mortgage payment households have higher liquid wealth, which could lead to the effective discount rate being closer to the long-term exponential discount rate. When looking at the drop in consumption at the same point in time for the pay-cycles with and without a mortgage payment, the greater drop can be caused by a steeper decline in consumption in the pay-cycles with a mortgage payment because of the lower effective discount rate.

1.4.3 Mental Accounting

In models of mental accounting, households use rules-of-thumb to determine their consumption paths.[27] Shefrin and Thaler (1988) distinguish between three different types of accounts – current spendable income, current assets and future income. While classic life-cycle models would treat all accounts as fungible, a household following a model of mental accounting may have a different propensity to consume depending on the account.

In our context, the sensitivity to the timing of a mortgage payment is consistent with a model of mental accounting in which households determine their current disposable income based on their bank account balance. When households repay their mortgage, which may be a particularly high cash outflow, their bank account balance decreases which in turn impacts household consumption. Because households are using their account balance as an indicator for their current disposable income it is the timing of changes in their bank account balance that drives changes in consumption.

Two findings support the possibility that household sensitivity to mortgage timing

[26] Kaplan *et al.* (2014) discuss the possibility that even wealthy households are liquidity constrained through holding mostly illiquid assets.

[27] Mental accounting models are often driven by empirical observations and typically do not form the basis of model with generalizable predictions. For an attempt to use mental accounting in a setup that nests other household finance models, such as the buffer-stock model and hand-to-mouth consumption, see McDowall (2019).

payment is driven by a mental accounting model. Firstly, our finding that households reduce debit card spending and not credit card spending suggests that households respond to bank account balance. The choice to cut back on debit card spending, which has a direct impact on bank account balance, and not credit card spending is particularly surprising given the number of households that pay interest on their credit cards. This surprising result possibly supports the notion that consumers are responding to bank account balance in way that translates directly into their bank account balances, which is true for debit cards but not credit cards.

An analysis of consumption by bank account balance tercile shows a greater consumption response to mortgage payment for households with lower bank account balances. Table 1.5 reports estimates from regression 1.1, similar to Table 1.2 but broken down by bank account balance tercile. For the households with the lowest bank account balances, we see a decline in average daily consumption of roughly 0.11 AUD in the week following a mortgage payment. For households in the highest bank account balance tercile the change in average daily consumption in the week following a mortgage repayment is indistinguishable from zero.

Differences in consumption decline based on household bank account balance tercile is distinct from other measures of household financial resources. Appendix Table A.2 reports changes in consumption by income terciles, Appendix Table A.3 reports changes in consumption by repayment amount terciles and Appendix Table A.4 reports changes in consumption by ratio of income to principal repayment amount. While in general separating our sample into terciles increases the noise of our estimates, the Appendix tables suggest there is no clear indication that lower income households or households for which their mortgage payment is likely to be more impactful respond more strongly to mortgage repayment.[28] The response specifically among households with lower bank account balances

[28] Lower income or a lower ration of income to mortgage repayment suggests that these households have less disposable income. However, classic models with liquidity constraints would still not predict a consumption response to the timing of mortgage repayment given that even households with less resources would save in anticipation of an outflow.

Table 1.5: *Daily Spending on Food, Drink and Entertainment Away from Home as a Function of Number of Days from Nearest Mortgage Payment, by Bank Account Tercile*

Panel A. Lowest Tercile

Sample	Full Sample			Households that have both debit and credit cards		
Cards included in spending measure	Debit and credit	Debit cards only	Credit cards only	Debit and credit	Debit cards only	Credit cards only
Days in mortgage repayment cycle						
-15 through -8	-0.006 (0.023)	-0.001 (0.012)	-0.004 (0.018)	-0.045 (0.034)	-0.014 (0.019)	-0.023 (0.026)
-7 through -1	--	--	--	--	--	--
0 through +6	-0.105*** (0.023)	-0.075*** (0.011)	-0.024 (0.018)	-0.144*** (0.035)	-0.107*** (0.020)	-0.025 (0.026)
+7 through +15	-0.078** (0.024)	-0.046*** (0.012)	-0.026 (0.018)	-0.083* (0.035)	-0.070*** (0.019)	-0.004 (0.026)
Number of households	15,534	15,534	15,534	7,385	7,385	7,385
Number of observations	7,808,612	7,808,612	7,808,612	3,794,931	3,794,931	3,794,931

Panel B. Mid Tercile

Sample	Full Sample			Households that have both debit and credit cards		
Cards included in spending measure	Debit and credit	Debit cards only	Credit cards only	Debit and credit	Debit cards only	Credit cards only
Days in mortgage repayment cycle						
-15 through -8	-0.047+ (0.026)	-0.016 (0.012)	-0.027 (0.020)	-0.081* (0.038)	-0.042* (0.019)	-0.031 (0.028)
-7 through -1	--	--	--	--	--	--
0 through +6	-0.070* (0.027)	-0.051*** (0.012)	-0.004 (0.021)	-0.104** (0.038)	-0.087*** (0.020)	0.003 (0.028)
+7 through +15	-0.057* (0.026)	-0.027* (0.012)	-0.025 (0.020)	-0.056 (0.038)	-0.049* (0.020)	-0.003 (0.030)
Number of households	15,534	15,534	15,534	7,720	7,720	7,720
Number of observations	7,658,143	7,658,143	7,658,143	3,912,598	3,912,598	3,912,598

Panel C. Highest Tercile

Sample	Full Sample			Households that have both debit and credit cards		
Cards included in spending measure	Debit and credit	Debit cards only	Credit cards only	Debit and credit	Debit cards only	Credit cards only
Days in mortgage repayment cycle						
-15 through -8	0.007 (0.027)	-0.016 (0.010)	0.021 (0.023)	-0.021 (0.039)	-0.029+ (0.016)	0.011 (0.032)
-7 through -1	--	--	--	--	--	--
0 through +6	-0.015 (0.027)	-0.015 (0.010)	0.000 (0.023)	-0.003 (0.040)	-0.019 (0.016)	0.015 (0.033)
+7 through +15	0.011 (0.027)	-0.005 (0.010)	0.016 (0.023)	-0.008 (0.039)	-0.008 (0.016)	0.006 (0.032)
Number of households	15,534	15,534	15,534	7,786	7,786	7,786
Number of observations	7,699,825	7,699,825	7,699,825	3,973,169	3,973,169	3,973,169

+ Significant at 10% level. * Significant at 5% level. ** Significant at 1% level. *** Significant at 0.1% level.

suggests it is the bank account balance to which some households are responding rather than their overall resources.

Chapter 2

Incomplete Contracts and Future Data Usage

Abstract

Most major jurisdictions require websites to provide customers with privacy policies. While privacy policies' most important function is to provide consumers with a description of online service providers' current privacy practices, we argue that these policies also serve a second, often-overlooked function: They allocate among online services and consumers the power to decide whether a service can modify its privacy practices and use consumer data in novel ways. We furthermore argue that a central feature of the E.U.'s General Data Protection Regulation (GDPR), one of the most comprehensive and far-reaching privacy regulatory regimes, is to restrict privacy policies that allocate broad rights for future data usage to service providers. We provide a theoretical explanation for this type of regulatory intervention by adapting standard models of incomplete contracts to privacy policies. We then use the model to explain how U.S. firms reacted to the GDPR. We show that U.S. websites with E.U. exposure are more likely to change their U.S. privacy policies to have *less* stringent and more lenient modification rules. Among websites that do not have E.U. exposure, we see the opposite trend. These results suggest that websites sought to

obtain residual rights over data usage after learning of the impact of the GDPR's stricter requirements.

2.1 Introduction

Privacy policies are ubiquitous on the internet. While jurisdictions around the world differ in their approaches to regulating data privacy, they are united in requiring websites to make privacy policies available to consumers. The most important function of these policies is to provide consumers with a description of online service providers' current privacy practices. They describe the extent to which these providers collect user's data and how they subsequently use this information. Most major jurisdictions hold online service providers accountable for breaches of their own privacy policies, rendering these policies a functional equivalent of consumer contracts governing other parts of the relationship between online service providers and the consumers they serve.[1]

In this paper, we argue that an important and often-overlooked feature of privacy policies is that they allocate among online services and consumers the power to decide whether a service can change its privacy practices and use consumer data in novel ways. This power is relevant, for example, if a provider considers rolling out new product features that require the collection or processing of consumer data in ways that are different from prior data usages. As an example, consider a social network deciding whether to introduce facial recognition technology to identify individuals in pictures uploaded by its users.[2] Will the introduction of this feature require a modification of the service provider's current privacy policy? If yes, does the consumer need to agree to this modification? The answer to both questions will, to a large extent, depend on the language of the current privacy policy.

We furthermore argue that a central feature of the E.U.'s General Data Protection Regulation (GDPR) and similar data privacy regulations is to protect consumers against

[1]There is an ongoing debate in the legal literature about whether courts should treat these policies as contracts. Some scholars argue that enforcing privacy polices would provide consumers with necessary protections, particularly with respect to data breaches Matwyshyn (2013). Others are skeptical about a contractual approach to privacy polices Karnow (2021). A similar disagreement exists over whether courts are *de facto* treating privacy policies as contracts. Some argue that courts tend to view privacy policies as "mere notices of company practice" Solove and Schwartz (2019). By contrast, a survey of privacy disputes that reached court judgements from 2016 suggests that in most cases the courts treated privacy policies as contracts Bar-Gill *et al.* (2017); Shahar and Strahilevitz (2016).

[2]Facebook has used facial recognition technology in this way since 2010.

privacy policies that allocate broad rights for future data usage to service providers. The GDPR requires privacy policies to explicitly describe the various ways in which service providers collect and use data. In other words, the GDPR contains a prohibition against general clauses that permit broad ranges of uses. In addition, while the GDPR allows for changes to a service provider's privacy policy, every new data collection or use requires a legal justification to be permissible. One potential justification is consumer consent. This means that a service provider cannot reserve for itself the right to unilaterally enact meaningful changes to its privacy policy.

The GDPR places few substantive limitations on the content of privacy policies and focuses primarily on what constitutes adequate notice and consumer agreement to policy terms. This is despite the growing recognition that privacy harms can occur outside the relationship between websites and users Viljoen (2021), raising questions about whether the fact that consumers voluntarily engage in these interactions is enough to guarantee socially optimal outcomes. In this paper we consider privacy policies through the predominant regulatory lens of bilateral exchange between websites and users, although we recognize that this is only a partial consideration of privacy harms, which is complex and multi-faceted.

Yet the GDPR's approach to future data usage rights is puzzling even when privacy policies are considered through a narrow bilateral lens of an exchange between websites and users. The GDPR's limits on broad allocations of rights to future data usage to service providers are difficult to reconcile with either *rational* or *behavioral* standard accounts of consumer conduct. According to rational accounts of consumer behavior, firms and consumers have necessary information and rational beliefs, and maximize utility (or profit) accordingly. Under this view, if a social network adopts a feature that consumers consider privacy-invasive, consumers who continue to use the service do so because they value it in spite of this feature Lenard and Rubin (2010); Westin (2000).[3] The GDPR's approach

[3]This is sometimes referred to as the 'privacy calculus' Acquisti *et al.* (2016). Early law and economics thinking on privacy supported the notion that privacy regulation can interfere with efficient markets, suggesting that consumers and firms benefit from minimal privacy restrictions Posner (1981); Stigler (1980). A focus in the recent years of privacy economics is to consider the various frictions in data markets that suggest a role of regulatory interventions, relaying on the existence of transaction costs and externalities that complicate a simple

does not seem compatible with this account of consumer behavior. If consumers are able to make meaningful tradeoffs between privacy harm and the benefits of using a website, there is no need for regulation to limit the (mutually beneficial) agreements between users and websites. Consumers and firms should be able to determine the scope of data usage rights at the outset, as well as the modification rules for policy changes.

According to *behavioral* accounts of consumer conduct, consumers are unable to detect when a firm makes adopts harmful privacy practices. Consumers by and large do not read privacy policies Obar and Oeldorf-Hirsch (2020),[4] and even if they did, it would be unlikely they would understand the policies and their implications Strahilevitz and Kugler (2016). Therefore, they do not consider harmful privacy practices when deciding whether to use a service. If consumers do not read or understand privacy policies, and merely consent policies they are presented with, the GDPR's restrictions on broad allocations of rights to future data usage to service providers are also not likely to be meaningful. Requiring consumers to agree to future policy changes does not solve the problem that consumers are likely to agree to policies without reading or considering their content. Therefore, the approach of the GDPR, perhaps the most influential data governance regulation worldwide, is hard to justify also on behavioral grounds.

This paper makes two contributions, one **theoretical** and one **empirical**. On the **theory** front, we explain how privacy policies allocation residual data usage rights and why a regulatory model like the one adopted by the GDPR can protect consumers from harm stemming from the allocation of broad rights over future data usage to online services. For this, we develop a theoretical framework based on standard models of incomplete contracts in Hart *et al.* (1997).[5] We note that the characterization of privacy policy as incomplete

---

story of trade-offs between consumers and firms (e.g., Romanosky and Acquisti (2009)).

[4]It is likely rational for consumers not to read privacy policies. A study from 2008 estimated that it would take 244 hours a year to read policies McDonald and Cranor (2008). According to a recent estimate it would take an average person 76 working days reading privacy policies people agree to in a year. "Amazon's terms and conditions along out loud takes approximately nine hours," N.Y. Times (Feb. 2, 2019), `https://www.nytimes.com/2019/02/02/opinion/internet-facebook-google-consent.html`.

[5]To our knowledge, the only other paper to apply the incomplete contracts theory to privacy is Dosis

contracts is functional rather than legal. In particular, this modeling choice does not require that privacy policies are treated as contracts in the legal sense.

Our more nuanced account of negotiations over privacy practices assumes that there are too many future contingencies for it to be possible to contract for all future states of the world at the point of time in which parties enter the initial contract.This means that parties determine at the time of contracting whether the website has a broad right to use consumer data, even for unspecified uses ('residual data rights'). If consumers retain the residual rights to their data, yet future developments or investments require changes in consumer data usage, websites must obtain consumer consent, which we consider a 'take-it-or-leave-it' renegotiation. The assumption of contractual incompleteness is not hard to motivate once one recognizes that the type of data collected and how it will be used cannot be fully specified at the initial point in time that the consumer agrees to a privacy policy. When college students first signed up for Facebook in 2005, it was not obvious that the company would one day be able to track their precise geolocation at all times, nor compare their facial structures in an automated way to a vast database of biometric data. The ability to collect consumer data and the ways in which firms can profit from this information depend on technological and business innovations that might not be known at the point of initial contract. Firms are aware that consumer information is valuable in general, but they only discover over time which consumer data they need to use to generate profits and how this data needs to processed to achieve this end.

This framework also provides insights into why privacy policies use open-ended clauses that allow websites broad future discretion in how to use consumer data rather than specifying usages. In the original incomplete contracts framework, parties allocate property rights that determine who has the residual rights over an asset. Although privacy policies rarely explicitly allocate data property rights,[6] they often use open-ended language to

---

and Sand-Zantman (2019), who study a setting in which a monopolistic firm offers a service which generates valuable data. Similar to our setting, the use of data is profitable for the firm but generates privacy costs for consumers. However, their analysis centers on a setting of heterogeneous users with respect to their value of the service and the firm looks to extract data while potentially distorting the service usage of high types.

[6]There is an extensive literature discussing whether there is data property right and whether there should

establish whether consumer data can be used in a way not explicitly stipulated in the policy. This data right determines whether a website requires consumer consent for any changes in data usage not explicitly mentioned in the policy, which we refer to as a 'renegotiation' as it adjusts the terms of the agreement.[7] The framework also provides predictions as to when it might be beneficial for the consumer to retain residual rights to their data.

We further build on the incomplete contracts framework by developing a *behavioral incomplete contracts* framework. In this model, a behavioral agent discounts future privacy harms at the point in time in which they enter the contract. This assumption is appealing because of how it corresponds to consumer understanding in the real world. Although consumers have some, albeit limited, understanding of the possible privacy harms caused by the current use of their data, they are likely to have a particularly weak understanding of how future innovations may affect privacy. In the model, consumers in some cases give the website residual data rights at the time of contracting when it would have been beneficial for them to retain the right and renegotiate with the website when the new uses and their harms are better understood.

We explicitly solve the model for the special case in which the firm has all market power and demonstrate that consumers may be indifferent, *ex ante*, between the firm or consumer holding residual data rights even when the firm holding residual data rights leads to a negative payoff. This is not the case when the consumer retains residual data rights and renegotiates only after they have rational privacy harm beliefs. Therefore, an important conclusion from the behavioral incomplete contracts model is that regulation has a role in intervening at the initial point of allocation of residual data rights. By limiting the ability

be a data property right (Bergelson (2003); Litman (2000); Victor (2013)). We do not directly address the question of whether data can be conceived of as property and instead focus on the ability to use data in ways that are not explicitly specified in privacy policies, which we consider a residual data right.

[7]We refer to the consent to a change in terms as a renegotiation, following the language used in many incomplete contract models, although in reality there is obviously no real negotiation between firms and consumers. In some loose way, however, firms may have to offer consumers a return to their consent for changing terms. For example, if Facebook wishes consumers to continue using their platform and consent to policy changes, it must continue to maintain and invest in their platform. In this way, the adjustments made by both the firm and consumer can be referred to as a renegotiation of the initial terms.

for data rights to be assigned to websites at the stage of initial contracting, firms will be required to renegotiate with consumers when they wish to implement a new innovation and when the privacy harms are better known to the consumer.

The *behavioral incomplete contracts* model provides an explanation for regulatory models like the one adopted in the GDPR. One way to understand the GDPR's requirements regarding explicit descriptions of services' data practices is that they limit the ability to assign the website the residual rights to data usage in the case of an incomplete contract. As future uses of data that depend on technological or business innovation cannot to be explicitly described at the initial point in time a consumer agrees to a policy, the GDPR blocks the possibility that consumers agree to these future uses. In other words, the consumer must retain residual data rights, requiring a renegotiation in the form of active consent if websites wish to change data collection and usage provisions. In essence, the GDPR is creating a two-part mandate for privacy policies, requiring (1) that residual rights over data usage be assigned to consumers and (2) a heightened modification rule.

The *behavioral incomplete contracts* model also sheds light on what the unconstrained choice of websites, when not exposed to the GDPR, might look like. The GDPR's requirement that consumers consent to changes to the policy, while limiting the ability to use broad language in the original policy, creates frictions and could be a meaningful constraint on websites. Therefore, when websites are unconstrained in their U.S. policies, they are likely to choose to allocate residual data rights to themselves knowing that consumers discount the full implications of allocating residual rights to the firm.

The **empirical** section of the paper informs the theory section by shedding light on whether there are frictions created by consumers retaining residual data rights (what we refer to as 'renegotiation' in the model) and whether firms seek to obtain residual data rights when unconstrained by regulation. Our empirical analysis is not a direct test of the theory but rather a close consideration of one aspect of the theory. Namely, and examination of privacy policy provisions that affect how privacy policies can be modified, and specifically, whether privacy policies stipulate that consumers must consent to policy changes. While not a formal test of our theory, our empirical results are consistent with firm behavior in our model.

In our analysis, we consider only privacy policies used in interactions with U.S. consumers. In these interactions, U.S. online services do not have to comply with the GDPR. Hence, we consider them unconstrained by European regulation: They are able to use open-ended language with respect to data usage; and they are also not required to conform with the GDPR's rules for policy changes. Within our sample of U.S. privacy policies, we compare websites that (also) cater to E.U. users, meaning that they must comply with the GDPR in some of their interactions with consumers, with websites that do not cater to E.U. users and therefore do not comply with the GDPR in any of their interactions. For every website, we compare privacy policies before and after the GDPR came into effect.

We hypothesize that the GDPR, through creating strict consent requirements, can lead websites to learn about the implications of requiring active consent for policy changes. Websites' experiences with the GDPR might translate into what they choose to do in their interactions with U.S. consumers. If websites with E.U. exposure learn that the GDPR consent requirement is costly, they may chose to limit their requirement to obtain consumer consent for changes in data usage in their interactions with U.S. consumers. By contrast, we do not expect a similar effect for website providers that were unexposed to the GDPR. While this analysis is not a direct test of the paper's theory, it can provide circumstantial evidence for the proposition that residual data usage rights matter, and that the GDPR affected service providers' allocation these rights.

We show that U.S. websites with E.U. exposure are more likely to change their U.S. privacy policies to have *less* stringent and more lenient consent requirements for policy changes. In 2018, before the GDPR came into effect, most U.S. privacy policies in our sample do not mention any consent requirement with respect to changes to the policy and continue to not mention any consent requirement in their 2020 policies. However, of the policies that chose to change their consent requirements, the E.U. exposed websites are more likely to have their later policy contain *less* stringent consent requirements, either by no longer requiring that consumers actively consent to changes or by dropping any mention of consent. Among websites that do not have E.U. exposure we see the opposite trend. Websites that do not have any E.U. exposure are more likely to add a consent requirement for changes or *increase* the stringency of the consent requirement.

Our results support the theory that the E.U. consent requirement, a central aspect of the restriction on allocation of residual data rights to websites, can be costly for websites, and that it is primarily E.U. exposed websites that are learning about the costly requirement. Websites with E.U. exposure then choose to adjust their U.S. policies, which do not need to comply with the GDPR, to avoid these costs by providing a weak requirement for consent, if at all.

These results contribute to our understanding of how regulation in one jurisdiction may affect firm behavior in another jurisdiction. The 'Brussels Effect,' a term coined by Anu Bradford (Bradford (2020)) refers to a situation in which the E.U. unilaterally regulates through voluntary firm compliance with E.U. regulations. This paper seeks to build on this framework of voluntary compliance through analyzing the response of firms to the GDPR in a particular context. Namely, consent requirements for privacy policy changes. Our results suggest that in this domain there is need for more direct regulatory intervention to shape firm practice with respect to privacy policy provisions on data usage. Understanding the theoretical underpinnings of the GDPR and its impact on website behavior help shed light on broader policy debates around restricting the substance of privacy policies and creating mandatory rules for policy modification.

Our empirical analysis also contributes to the literature on the effect of the GDPR on consumers and firms in the E.U. Some studies consider the degree to which privacy policies actually comply with the GDPR, with mixed results (Kretschmer *et al.* (2021); Lancieri (2022)). Other studies have focused on consumer behavior in response to the GDPR (Santos *et al.* (2021)). Schmitt *et al.* (2021) demonstrate the impact of the GDPR on the number of website users, documenting a short term and long term decline in usage that varies across websites, with more popular websites suffering less.[8] Our results are consistent with these studies, as we suggest a new avenue in which compliance can be costly for websites, namely, the requirement to actively obtain consumer consent for changes in privacy policies. Our study differs from previous literature in that we are primarily focused on websites' unconstrained responses to the GDPR when they are not required to comply with the GDPR, providing a broader perspective on the effect of the GDPR outside the E.U. and firm preferences.

The paper has the following structure. In Section 2.2, we discuss the incomplete contract model of privacy policies on how it helps explain regulatory interventions like the GDPR and the ways in which websites respond. We build on the incomplete contract model by considering a case in which a behavioral agent mispredicts the harm resulting from the future innovations. In Section 2.3, we discuss our empirical study of U.S. privacy policies before and after the GDPR came into effect. Our main result is that websites that are also exposed to the GDPR are more likely to select a less stringent consent requirement after the GDPR by dropping any mention of consent to policy changes.

2.2 An Incomplete Contract Theory of Privacy Policies

To study the implications of allocating residual data usage rights, we consider a bilateral relationship between a firm and a mass of consumers. Consumers decide whether to use the service provided by the firm, while the firm decides the level of investment in an innovation that is profitable but can cause privacy harm to consumers.[9]

---

[8]Recent studies suggest a nuanced effect to some of the GDPR requirements. Aridor *et al.* (2020) show how the opt-in GDPR requirement leads to a drop in observable consumers but the remaining consumers are observable for longer. Mangini *et al.* (2020) attempt to more directly measure the costs of compliance with GDPR, primarily in the context of the right to be forgotten.

[9]Prior work on the theory of privacy focused on the economic implications of more information sharing versus more privacy Acquisti *et al.* (2016). The main conclusion from surveying the literature is that the welfare implications of more or less privacy depend on the particular context of information sharing and that "both individuals and organizations face complex, often ambiguous, and sometimes intangible trade-offs" (Acquisti *et al.* (2016), page 462). One approach to model internet privacy is where sellers search for buyers and send solicitations, while buyers (who have heterogeneous preferences for the seller's good) dislike being solicited and invest in hiding from the seller (Hann *et al.* (2008)). The authors argue that putting a price on solicitations increases welfare. The model does implicate some of the concerns of how new technology harms privacy, but it focuses on one specific type of privacy invasion, and ignores the type of privacy harm that we focus on and

We start by considering a firm's profit-maximizing level of investment in an innovation when the level of investment is fully contractible. In this case, the firm can commit to the level of innovation so that the level of investment does not depend on whether the firm or the consumer has the residual data usage rights. This provides a benchmark for the efficient level of investment. We then assume that the level of investment is not contractible and investigate the economic consequences of allocating the residual data usage rights to one of the sides of the market. Finally, we consider a case in which a behavioral consumer mispredicts the harm from future investments and only learns of the true harm after the firm selects the level of innovation. We show that in certain cases, this will lead to consumers giving up on their residual rights when it would have been socially optimal for them to retain these rights. The results of this model suggest that regulation has a role in intervening through limiting consumers' ability to transfer residual data rights to websites.

2.2.1 Model Assumptions

We consider a bilateral relationship between a web user or consumer, C, and a firm, F, who provides the consumer with a service. The consumer, C, receives some basic benefit B_0 in return for some basic price P_0.[10] The firm has some basic cost in providing the service to the user K_0. We also assume that the consumer has an outside option of zero. For example, consider a professional networking website, like LinkedIn,[11] that allows users to create pages with personal and professional information, which is then used to connect to other users and potential employers.

---

which is a bigger problem: how websites use personal data. While early writing focused narrowly on the relationship between the consumer and firm, in what Acquisti *et al.* (2016) characterize as the second wave of economics of privacy research, economists considered concerns created by the secondary usage of personal data, which was not considered in the first wave (see e.g. Varian (2002)). Another approach is a sequential contracting model in which an upstream seller gathers information about a buyer, which it can then sell to a downstream seller (Calzolari and Pavan (2006)). This model is potentially useful, especially in a dynamic setting. In this paper, we focus on a simpler, bilateral relationship that still captures the heart of the privacy debate.

[10]In many cases $P_0 = 0$. This is true for websites in which there is no upfront cost for using the websites services

[11]`https://www.linkedin.com`

Innovation.–The firm, F, can make a revenue increasing investment e that increases revenue by $r(e)$. The investment e is meant to capture a broad range of innovations, such as new technologies that allow consumer data to be collected, business innovations that create new revenue opportunities from consumer data and legal innovations that reduce the exposure of firms to litigation. For example, a professional networking website, like LinkedIn, might innovate with respect to the information it collects on users and then shares with recruiters.[12] Other innovations might include the sale of data to third party advertisers, operating on and off the the networking platform.[13] Innovations may also include business innovations, such as investing in business relationships and developing business models that would allow the platform to profit from consumer information. We assume that the cost of each unit of innovation for the firm is constant and equal to 1.

Costs.–Another key element of out model is that while the innovation creates value for the firm, it can be costly for consumers and causes harm to the consumer's privacy $h(e)$.[14] There are at least two relevant types of costs. The first cost is the psychological harm caused by the preference for privacy. While these costs are intangible they have been documented as important for consumers (Feri *et al.* (2016)).[15] Another type of cost is related to the economic impact of data sharing, namely the possibility of price discrimination when more information is available about consumers (Acquisti *et al.* (2016); Richards (2021)).[16]

[12]In fact, a significant source of LinkedIn's revenues comes from 'Talent Solutions' which allows recruiters access to information, ability to create targeted job posting and communicate directly with users.

[13]There are other more speculative types of innovations that LinkedIn could engage in. For example, LinkedIn could develop the ability to track other online behaviors which it can then sell to potential employers. See example of discussion of LinkedIn's revenue- `https://blog.waalaxy.com/en/how-linkedin-makes-money/`

[14]We assume that $r(e), h(e) > 0$. We also make the following assumptions in order to ensure a unique solution: $h(0) = 0; h' > 0; h'' > 0$. For the revenue function we assume that $r(0) = 0; r' > 0; r'(0) = \infty; r'(\infty) = 0; r'' < 0$. We also assume that $r' > h'$, which means that when a firm invests in an innovation, the marginal cost savings outweigh the marginal harm to the consumer.

[15]Understanding and estimating people's preference has been addressed in many papers in recent years (Acquisti *et al.*, 2016, section 3.8). One of the challenges for this literature is reconciling stated preferences for privacy and revealed choices, often known as the 'privacy paradox.'

[16]Another type of economic harm is the cost incurred when data is breached or misused, which may result in identity theft.

The benefits to the consumer and profits for the firm can be summarized as:

$$B = B_0 - P_0 - h(e)$$
$$R = P_0 - K_0 + r(e) - e$$

2.2.2 Contractible Benchmark

As a benchmark, we consider the firm's profit maximizing level of investment by assuming that this level of investment is contractible. This assumes that all future types of investments are known at time $t = 0$ when the parties enter the contract. We also assume that consumers make rational choices that balance privacy harms and benefits.

When investments are predictable and contractible, the consumer C and firm F choose the cost saving investment e to maximize surplus, solving:

$$\max_{e}\{-h(e) + r(e) - e\}$$

The optimal level of cost investment, expressed as the first order condition is:

$$-h'(e^{*}) + r'(e^{*}) = 1$$

The optimal level of investment depends on the marginal benefit of spending extra effort to reduce costs, measured to take account of the marginal harm to privacy, which must equal the marginal cost of that extra effort, equalling 1.

The surplus that results from the innovation can then be split between the firm and consumer according to their relative bargaining position at time $t = 0$.

2.2.3 Incomplete Contract

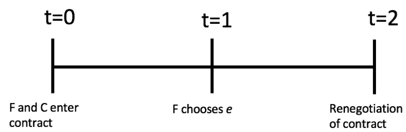
In an incomplete contract framework, we assume that there are too many contingencies that it is impossible to anticipate and contract for all future possibilities in advance. The harms and benefits that result from the investments are all observable but not verifiable to third parties, meaning that they are not contractible. These assumptions are at the core of the incomplete contract framework as they explain why contracts are necessarily incomplete, by not being specific enough to create certainty as to the rights and obligations in all future states. These assumptions seem particularly appropriate for the privacy setting.[17] The monetizing of consumer data depends on many future innovations regarding the processing and sharing of the data, such as the ability to store and analyze large datasets. Moreover, monetizing of consumer data can depend on future business innovations and intermediaries, such as third party that aggregate consumer data for marketers. Many of these future possibilities may be unknown or accessible at date $t = 0$, when the parties enter the contract.

Given the inability to specify all future data usages, parties may instead decide at the time they enter the initial agreement whether or not the firm has the right to use the data for future usages not explicitly agreed upon by the parties. We therefore consider the firm to have the residual right to data usage when it can use consumer data for usages not specified at the time the parties initially enter the contract. Conversely, the consumer has the residual right to data usage if the firm must obtain consumer consent for data usage not specified in the initial contract. This means that the firm would need to 'renegotiate' with the consumer for further uses of the data.

As a shorthand, we sometimes refer to this residual right to data usage as a quasi-property right, or refer to situations in which the firm 'owns' the data or consumers 'own' the data. This language closely relates to the broader literature on incomplete contracts that often focuses on the key allocation question of which party has the residual control or decision right (Grossman and Hart (1986)). However, it is important to note that whether data property rights can be assigned to firms or whether we should treat data as property all together, given its non-rivalrous nature, is a a topic that has been subject of much debate and is not fully addressed in this paper (Bergelson (2003); Litman (2000); Victor (2013)). Instead we assume that at $t = 0$ the parties enter the contract in which they assign the residual rights to the use of consumer data, which will then determine whether the firm can decide without the consent of the consumer to use the data for purposes not specified in

[17] We are aware of only one other paper that has used the incomplete contracts framework to study privacy. Dosis and Sand-Zantman (2019) consider the assigning of data property rights in a context of heterogeneous consumers, where the firm must incentivize consumer effort that generates monetizable data.

the contract.



Timeline.–At t=0 the website and user enter an agreement that assigns residual data right. At $t = 1$ the firm makes a decision with respect to their investment in a technology that will increase their profitability but also harm consumers. The implementation of the technology at $t = 2$ depends on who holds the residual rights to the data. If the consumer retains the residual right to the data, the firm will need consumer permission to implement the technology, requiring renegotiation with consumer. If the firm has the residual rights to the data, they are able to implement the technology without renegotiating with the consumer. We assume that all surplus gained from renegotiation is split in accordance to market power, where $\lambda \in [0, 1]$ notes the market power of the firm.

Renegotiation.–Referring to consumer consent to future data usages as 'renegotiation' is somewhat awkward. We use this terminology because it tracks how other incomplete contract applications refer to the stage at which the one party must obtain permission from the party with the residual datat rights as renegotiation. In reality, consumers and websites do not engage in a traditional negotiation in which consumers consent to data usage in return for a share of the surplus from the data usage. However, the take-it-or-leave offer presented by websites still requires that consumers perceive an ongoing benefit from using the website. For example, a website might know that in order to retain users in must continue to invest in the development and maintenance of it's interface and the service it provides to consumers. Absent this continuing investment, consumers may decide not to continue using a website and thereby refuse consent to data usage changes. Therefore, we can conceive of the interaction between users and websites as some loose form of renegotiation.

Firm has Residual Data Rights

If firm F holds the residual rights to the data, they are able to implement the cost investment without the consumer's permission. When the firm 'owns' the data, the party payoffs are:

$$U_c = B_0 - P_0 - h(e)$$
$$U_f = P_0 - K_0 + r(e) - e$$

The firm will select the investment level that maximizes its utility, so that e_f expressed by the first order condition is:

$$r'(e_f) = 1$$

The total surplus is:

$$S_f = U_f + U_c = B_0 - K_0 - h(e_f) + r(e_f) - e_f$$

P_0 is chosen to allocate this surplus between the parties according to their relative bargaining positions at time $t = 0$.

We can compare the level of innovation under a complete contract framework and the incomplete contract framework, when the firm 'owns' the data:

	cost innovation
Complete contract	$-h'(e^*) + r'(e^*) = 1$
Firm owns data	$r'(e_f) = 1$

Under the incomplete contract framework, when the firm 'owns' consumer data they over-invest in cost innovation relative to a complete contract (which we can consider as the first best), so that $e_f > e^*$. This is because the firm does not internalize the deterioration in consumer privacy that results from the innovation, and instead invests in cost innovation until the marginal cost saving equals the marginal cost (which is one).

Consumer has Residual Data Rights

When the consumer 'owns' the data, the firm is unable to implement an innovation without the permission of the consumer. This means that, to implement the cost innovation, the parties must renegotiate. We assume again that the surplus from the renegotiation is split according to the relative market power of the firm, expressed by λ.

When the consumer 'owns' the data, the parties' payoffs are:

$$U_c = B_0 - P_0 + (1-\lambda)\left[-h(e) + r(e)\right]$$
$$U_f = P_0 - K_0 + \lambda\left[-h(e) + r(e)\right] - e$$

This assumes that the cost of the innovation is not included in the calculation of surplus.

The firm will select the investment level at $t = 1$ that maximizes its utility, so that e_c expressed as the first order condition is:

$$\lambda(-h'(e_c) + r'(e_c)) = 1$$

The total surplus is:

$$S_c = U_c + U_f = B_0 - K_0 - h(e_c) + r(e_c) - e_c$$

P_0 is chosen to allocate this surplus between the parties according to their relative bargaining positions at time $t = 0$.

We can compare the cost innovation the under complete contract framework and the incomplete contract framework, when the firm or consumer own the data:

	cost innovation
Complete contract	$-h'(e^*) + r'(e^*) = 1$
Firm owns data	$r'(e_f) = 1$
Consumer owns data	$\lambda(-h'(e_c) + r'(e_c)) = 1$

We can summarize that the level of innovation is higher when the firm owns the data than when the consumer owns the data $e_f > e_c$. This is because the firm internalizes some of the harm from the cost innovation when the consumer owns the data. When the consumer owns the data, the innovation level is lower relative to its first best level in the contractible case, so that $e^* > e_c$. This is because the firm only internalizes part of the net benefit of the cost innovation, as opposed to the complete contract case in which the firm internalizes the full net benefit of the cost saving innovation. The conclusion is that allocating residual data usage rights to the firm and the consumer suffers from an inefficiency relative to the first best. When the firms owns the data they overinvest in innovation and when the consumers owns the data the firm underinvests in innovation.

Efficient Allocation of Residual Data Rights

The allocation of residual data rights to the firm is efficient if $S_f > S_c$.

$$S_f > S_c \iff -h(e_f) + r(e_f) - e_f > -h(e_c) + r(e_c) - e_c$$

Conversely, the allocation of residual data rights to the consumer is efficient if $S_f < S_c$.

In allocating residual data rights, the firm and consumer trade-off the inefficiency of over-investment (when the firm has residual rights) and under-investment (when the consumer has residual rights). Therefore, efficient allocation depends on the comparison of the harm function $h(e)$ and the revenue function $r(e)$. When the harm function is not significant allocation of residual data right to the firm is superior. This is because when the harm from the cost innovation becomes small, the cost investment under firm's holding residual data rights is closer to the first best. Conversely, if the privacy harm is not adequately offset by the benefit, allocation of residual data rights to consumers is superior. In such a case it is more efficient to have lower incentives to invest in the cost innovation, like under consumer residual data rights.

A key prediction of the model, therefore, is that if firms and consumers allocate data rights efficiently, allocation of residual data rights to firms is more likely when the harm to privacy is less severe or when the benefits of the technology offset the harms. An example of when data innovation is very valuable to the firm but does not cause much harm to privacy, is when the firm makes use of anonymous data that can be used to train an algorithm. In that case it could be more efficient for the firm to have residual data rights.

Similarly, we would expect allocation of residual rights to consumers when the benefit to firm is more likely to be offset by the harm. For example, a dating app providing information on people's dating behavior to a third party. It is plausible that the harm is close to the benefit. More generally, the more sensitive the data that the firm is privy to, the more efficient it is for the consumer to have residual data rights.[18]

2.2.4 Behavioral Incomplete Contracts

In the incomplete contracts model both parties are rational but are unable to achieve the first best level of investment because they are unable to contract on the optimal level at the time of the initial contract. In this section we assume that the consumer is not fully rational. Specifically, we assume that the consumer, at the time they enter the contract, is naive about future privacy harm of innovation and so that they believe the harm function is $\delta h(e)$ rather than $h(e)$, where $0 < \delta < 1$. At $t = 0$ the consumer assumes that the firm will select the level of investment at $t = 1$ that depends on the cost function of $\delta h(e)$. Unlike the consumer, the firm assumes the correct cost function $h(e)$. At $t = 1$, after the firm has made an investment decision, the true hard function becomes known to the consumer. If the consumer 'owns' the data the parties then renegotiate when both parties accurately understand the privacy harm of innovation.

Firm has Residual Data Rights

If the firm owns the data, there is no negotiation at $t = 2$. At $t = 1$ the firm selects its level of investment to maximize its utility $U_f = P_0 - K_0 + r(e) - e$.[19] Therefore its level of investment, expressed as a first order condition is:

$$r'(e_f) = 1$$

[18] A high profile example of a data breach that caused a significant amount of harm due to the sensitivity of the data collected is the data breach of the website Ashley Madison. See `https://en.wikipedia.org/wiki/Ashley_Madison_data_breach`.

[19] Although P_0 depends on the level of innovation e, as will be explained, at the time the firm makes the decision of the optimal level of innovation this has no effect on the price paid by the consumer in at $t = 1$.

The consumer believes (incorrectly) at $t = 0$ that its utility under allocation of residual rights to the firm will be:

$$U_c^b = B_0 - P_0 - \delta h(e_f)$$

When in fact its utility at $t = 2$ will be:

$$U_c = B_0 - P_0 - h(e_f)$$

$U_c^b > U_c$ and so the consumer's utility is lower than what she predicted it would be. This is because the consumer discounts the harm that will result from the firm holding residual data rights. If the basic price to access the website, P_0 is set endogenously[20] the behavioral consumer may be willing to pay a higher basic price so that $P_0^b > P_0$. This would provide a clear benefit for the website, as they are able to charge a higher price as a result of the consumer's misprediction as to future innovation levels or use the lower price to induce the consumer to agree to firm ownership when the firm prefers firm ownership.

Consumer has Residual Data Rights

To understand how the behavioral consumer's decision and utility differ from the rational consumer's decision and utility, we must first understand the renegotiation at $t = 2$ and firm decision at $t = 1$.

t=2.– At $t = 2$ the parties renegotiate if the consumer retains the residual rights to the data. We assume that at $t = 2$ the consumer is aware of the **true** harm $h(e)$, and that the surplus from the negotiation is split according to the relative market power λ. This means that the party payoffs are:

$$U_c = B_0 - P_0 + (1 - \lambda)\left[-h(e) + r(e)\right]$$
$$U_f = P_0 - K_0 + \lambda\left[-h(e) + r(e)\right] - e$$

[20] In reality, the basic price of $P_0 = 0$ is quite sticky for many types of websites, particularly for basic access to the website.

The same as in the non-behavioral case.

t=1.– Realizing that the payoffs at $t = 2$ will be determined by the true harm function, the firm will select the investment level that maximizes its utility, so that e_c expressed as the first order condition is:

$$\lambda(-h'(e_c) + r'(e_c)) = 1$$

Again, the level of investment chosen by the firm is the same as in the non-behavioral case.

t=0.– From the consumer's perspective at $t = 1$ they have incorrect beliefs as to the level of investment that the firm will select at $t = 1$. The consumer believes that the payoffs at $t = 2$ will be:

$$U_c^b = B_0 - P_0 + (1 - \lambda)\left[-\delta h(e) + r(e)\right]$$
$$U_f^b = P_0 - K_0 + \lambda\left[-\delta h(e) + r(e)\right] - e$$

So that the consumer believes that the firm at $t = 1$ will select an investment level that maximizes what the consumer expects to be the firms utility:

$$\lambda(-\delta h'(e_c^b) + r'(e_c^b)) = 1$$

Note that $r(e_c^b) > r(e_c)$, meaning that the consumer believes that the firm will chose a higher level of investment than what the firm actually ends up selecting in $t = 1$. If we compare these levels of investment to the level of investment under firm residual rights we get that $r(e_f) > r(e_c^b) > r(e_c)$, meaning that the level of investment is higher when the firm holds residual data rights than under the consumer's incorrect beliefs at $t = 0$ when consumer retains residual data rights, which is higher than the true level of investment under consumer residual rights.

Recall that in Section 2.2.3 we discussed that whether it was optimal for the consumer or the firm to have residual rights to data depended on the relative benefit and harm from the innovation. Firm's having residual data rights is more likely when the benefit of the innovation offsets the harm from the innovation. Therefore, in the behavioral case:

1. When the firm holds residual rights, the consumer perceives the harm to be lower than the true harm of the innovation, meaning that the behavioral consumer is more likely to believe the benefit of the innovation offsets the harm of the innovation.
2. However, when consumer's retain residual rights, the consumer believes the level of investment will be higher than it really is. Therefore the consumer believes that the difference between the level of investment with firm residual rights and consumer residual rights is smaller than it truly is.

The first factor leads to the consumer discounting the harm of firm ownership while the second factors mitigates concerns over consumer ownership. From the firm's perspective, the discounted harm at $t = 0$ leads to the a higher to the consumer paying a higher price for the initial access to the service. The next section considers these factors under the assumption that the firm has all market power.

Special Case: Firm has all Market Power

To understand the implications of the behavioral incomplete contracts model, I consider a special case in which the firm has all market power both at the time the parties enter the initial contract, $t = 0$, and when the parties renegotiate at $t = 2$. If the firm has all market power at $t = 0$, and we assume that the consumer has a reservation price of 0, the consumer will only enter the contract if the expected payoff is non-negative. If the firm has all market power at $t = 2$ we set $\lambda = 1$.

Firm has residual data rights.–Although the consumer believes the future harm function to be $\delta h(e)$, the consumer has rational beliefs about the level of investment selected by the firm because the firm does not internalize the privacy harm under firm ownership. The firm selects e_f, the same as the non-behavioral case.

This means that the price P_0 is determined by setting the consumer payoff to 0

$$U_c = B_0 - P_0^{bf} - \delta h(e_f) = 0$$

$$P_0^{bf} = B_0 - \delta h(e_f)$$

Relative to the maximum price under rational beliefs, the consumer is willing to pay a higher price because they discount future harm- $P_0^f < P_0^{bf}$.

The firm payoff is:

$$U_f = P_0^{bf} - K_0 + r(e_f) - e_f$$

$$U_f = B_0 - K_0 + r(e_f) - e_f - \delta h(e_f)$$

Consumer has residual data rights.–When the firm has all market power at $t = 2$ it is able to fully capture the surplus from renegotiation (but still needs the consumer's consent for implementing the innovation and therefore must compensate the consumer for the privacy harm for the consumer to agree). This means that the firm maximizes:

$$\max_e \{-h(e) + r(e) - e\}$$

The optimal level of cost investment, expressed as the first order condition is:

$$-h'(e^*) + r'(e^*) = 1$$

This is the same level of investment as under the contractible benchmark.

However, because the consumer believes that the harm is onlyδ the consumer believes that the firm is maximizing:

$$\max_e \{-\delta h(e) + r(e) - e\}$$

The optimal level of cost investment, expressed as the first order condition is:

$$-\delta h'(e^{bc}) + r'(e^{bc}) = 1$$

The consumer believes that the level of investment selected by the firm is higher than e^*. This is because the harm from the innovation is lower. The consumer expects to withhold their agreement to the implementation of the innovation unless they are compensated for the privacy harm $h(e^{bc})$.

The initial price the consumer pays for the service (P_0^{bc}) is determined by setting the consumer's believed payoff to 0:

$$U_c = B_0 - P_0^{bc} - \delta h(e^{bc}) + \delta h(e^{bc}) = 0$$

$$P_0^{bc} = B_0$$

The maximum price the consumer is willing to pay for access to the service is higher than under firm ownership. The consumer beliefs about the level of investment is lower under consumer ownership ($e_f > e^{bc}$) because under firm ownership the firm does not internalize the privacy harm at all.

What actually happens is different to the consumer's beliefs. When the parties renegotiate at $t = 2$ the firm must adequately compensate the consumer for the harm of the innovation, because otherwise the consumer would not agree to the implementation of the innovation. This means that the firm actually compensates the consumer by $h(e^*)$ and not $\delta h(e^{bc})$

The firm payoff is:

$$U_f = P_0^{bc} - K_0 + r(e^*) - e^* - h(e*)$$

$$U_f = B_0 - K_0 - h(e^*) + r(e^*) - e^*$$

When the firm prefers to have the residual data rights.–The firm prefers firm ownership when its payoff under firm ownership is higher than its payoff under consumer ownership.

Firm payoff under firm ownership is:

$$U_f = B_0 - K_0 + r(e_f) - e_f - \delta h(e_f)$$

Firm payoff under consumer ownership is:

$$U_f = B_0 - K_0 - h(e^*) + r(e^*) - e^*$$

Therefore a firm prefers firm ownership when:

$$r(e_f) - e_f - \delta h(e_f) > r(e^*) - e^* - h(e^*)$$

Whether the firm prefers firm ownership depends on the consumer's beliefs δ and the harm and benefit functions.

Conclusion.–

1. For certain values of δ, and harm function $h(e)$ and benefit function $r(e)$ the firm may prefer firm ownership because under firm ownership the consumer pays a higher price for access to the service than they would with rational beliefs.

2. When the firm has all the market power, consumers with wrong beliefs about the harm function should not have entered an agreement for firm ownership altogether, as this results in a negative payoff, $U_c = B_0 - (B_0 - \delta h(e_f)) - h(e_f) < 0$

2.2.5 Model Implications

The goal of our model is to contribute to the conversation about privacy policies among scholars and policy makers. Specifically, our adaptation of Hart *et al.* (1997) provides insight into why privacy policies use open-ended clauses that allow websites broad future discretion in how to use consumer data rather than specifying usages. The model also demonstrates how allocation of residual data rights can affect the level of innovation and privacy protections. Our behavioral extension shows that when firms are better-placed than consumers to predict future privacy harms they may seek to obtain residual data rights or make future negotiations easier at the initial time of contracting. In our special case of when the firm has all market power, this could lead to firms retaining residual data rights even when this results in a transaction that the consumer would not have entered at all with rational beliefs. This conclusion implies that regulation may have a role in intervening at the initial point of allocation of residual data rights. By limiting the ability for data rights to be assigned to websites at the stage of initial contracting, firms will be required to renegotiate with consumers when they wish to implement a new innovation and when the privacy harms are better known to the consumer.

The regulatory implications of the behavioral incomplete contracts framework may capture the essence of the limitations set by the GDPR with respect to policy changes. The

GDPR's requirement that websites explicitly outline data usage in the privacy policy and the restrictions on policy modification can be seen as a method of forcing consumers to retain residual rights to their data. Under these types of regulatory interventions, firms are unable to hold residual rights to data usage and are therefore forced to renegotiate with consumers for any change in data usage. This renegotiation is typically in the form of a take-it-or-leave-it offer in which the consumer consents to the new data usage in return for the continued use of the website.

Our theory provides a framework to explain certain features of the GDPR, but it is unclear whether the assumptions of the model hold true in reality. We are therefore interested in understanding how websites respond to the GDPR and whether their response is consistent with the model. Specifically, we are interested in the unconstrained choice of websites that are exposed to the GDPR, meaning what websites choose to stipulate in the privacy policies they use for their U.S. consumers, where they are not subject to the GDPR. If the GDPR's limits on the use of broad language in privacy policies and its unalterable requirements for adopting changes to these policies create frictions and impose meaningful constraints on websites, websites might learn about the implications of consumers retaining residual data rights. Thus, in situations in which they are unconstrained by the GDPR (in interactions with U.S. consumers), they might be more likely to choose to allocate residual data rights to themselves. While this analysis is not a direct test of the paper's theory, it can provide circumstantial evidence for the proposition that residual data usage rights matter, and that the GDPR affected service providers' allocation these rights.

2.3 How do Websites Respond to the GDPR?

In this section, we study changes in privacy policies that affect how these policies can be changed. In particular, we are interested whether privacy policies stipulate that consumers must consent to future changes of the privacy policy. As we describe below, under the GDPR, websites are constrained with respect to the rules for changing privacy policies in two ways. Firstly, websites are required to explicitly lay down the types of consumer data that is collected and how this data is used, restricting website's ability to capture future changes in data collection and usage in their initial policy. Second, websites are required to justify any changes to their privacy policy that leads to more invasive privacy practices, which often requires obtaining consumer consent. The bottom line is that the GDPR lays down strict rules on consumer consent for future data rights.

In the U.S., websites do not face similar legal constraints. Therefore, we consider the unconstrained choices websites make in privacy policies used in interactions with U.S. consumers with respect to the provisions that govern changes in these policies. We hypothesize that the GDPR, through creating consent requirements, allows websites to learn about the implications of requiring active consent to policy changes. What websites learn will potentially translate into what they choose to do with their U.S. policies.

Not all U.S. websites may learn equally about the cost of consent for policies changes. If websites that are exposed to the GDPR, i.e. they must comply with the GDPR in their interactions with consumers in the E.U., they may be directly and concretely learning about these costs. On the other hand, for websites that do not cater to E.U. subjects, the costs of the consent requirement may be less salient and they might not be learning about the implications of the requirement.[21]

We consider whether websites learn about the costs of active consent to policy changes by comparing the privacy policies of GDPR-exposed websites and websites without GDPR exposure, before and after the GDPR went into effect. We show that GDPR-exposed websites are more likely to adopt less stringent rules with respect to changes in privacy policies.

[21]Note that our sample of websites does not include websites run by online service providers headquartered in the E.U., which are under a legal obligation to comply with the GDPR in their interactions with consumers worldwide (GDPR Art 3(1)).

2.3.1 Legal Landscape

General Observations

In both the U.S. and Europe, regulatory efforts have focused on a 'notice and consent' approach. Under this approach, online services are required to notify consumers about their privacy practices. Insofar as they follow the requirements for notice and consent in the respective jurisdiction, there are few—if any—substantial restrictions on what online services can do with consumer data.[22] Despite their common focus on notice and consent, however, the United States and the European Union take fundamentally different approaches to regulating data privacy. In the United States, data privacy regulation is mostly limited to specific fields such as education and credit reporting. Only in recent years have several states started to enact more comprehensive rules.[23] As a result of the restraint on the part of legislators and regulators, it is largely left to businesses and consumers to specify the rules governing the use of the consumer's data by way of contracting Davis and Marotta-Wurgler (2019).

Under E.U. law, any business that intends to collect or process personal data requires a legal justification to do so. One important way to secure this justification is by obtaining a consumer's consent (GDPR Art. 6(1)a.). Other avenues include showing that the "processing is necessary for the purposes of the [business's] legitimate interests" (GDPR Art 6(1)f.) Because of the relatively strict rules governing consumer consent, regulatory agencies and courts play a larger role in determining the scope of permissible data practices.

The E.U.'s approach to data privacy has been relatively consistent since at least the 1990s, when the first E.U.-wide data privacy law, the Data Protection Directive, entered into force. Since then, the most important milestone in the development of E.U. data privacy law was the entry into force of the GDPR in May 2018. Compared to the Data Protection Directive, the GDPR imposes substantially higher hurdles on businesses intent on collecting

[22]Sloan and Warner (2014) provides a summary of the notice and consent approach and its main critiques.

[23]The first of this new generation of privacy law was the California Consumer Privacy Act of 2018 (CCPA), which went into effect in January 2020.

and processing consumer data. For example, under the GDPR, a request for consent must "be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language" (GDPR Art. 7(2)).

A second important innovation of the GDPR is that it significantly expanded the geographic reach of E.U. data privacy law. While the GDPR's predecessor, the Data Protection Directive, applied only to businesses that were either operating out of the European Union or at least used equipment situated within the E.U. in their data processing (Data Protection Directive, Art. 4), the GDPR also applies to businesses located elsewhere, if and insofar these businesses 'target' consumers in the European Union (GDPR, Art. 3(2)).

Restrictions on Future Data Use

The different regulatory approaches in the United States and the E.U. also translate into different restrictions on future data usage.

In the E.U., businesses need to state ex ante in their privacy policies what type of data they collect and how they intend to use it. This was the case even before the GDPR entered into force. However, the GDPR significantly increased the level of detail with which businesses have to describe the use of consumer data. As a result, under the GDPR, it is impermissible for businesses to reserve the right to use data for purposes that were not explicitly mentioned in their privacy policies at the time of data collection. Instead, in many cases, the only way to use data gathered in the past in novel ways will to be to obtain a consumer's explicit consent for the new processing techniques.

In the United States, as described above, websites are generally free to use consumer data in any way that is not in conflict with their own privacy policy. At the same time, the law imposes few constraints on the contents of these privacy policies. Against this background, websites can *de facto* reserve the right to change their processing of consumer data after its collection by using broad terms that can accommodate a range of different data practices.

The extent to which services in the United States face legal constraints when changing their privacy policies is subject to debate. Given that there exists no comprehensive data privacy law in the United States, there are a number of legal grounds that could create constraints on changing policies.

One important set of constraints on changing policies follows from the threat of enforcement actions under Section 5 of the FTC Act. According to the FTC's enforcement practice, services are generally free to change their privacy policies at any time. One exception to this rule concerns retroactive changes. Retroactive changes are instances in which a service provider changes its privacy policy to treat data that has already been collected in a way that would have been in conflict with the policy that was in place at the time. Such changes are considered a violation of Section 5 of the FTC Act, unless the service provider obtains the "informed consent of affected consumers" (FTC in re: Gateway Learning Corp. 2004). In addition, to the extent that services lay down other constraints to changing their privacy policy, such as requiring consumer consent, they are not free to act in a manner inconsistent with the provisions in the privacy policy.

The ability of services to change their privacy policies might also be constrained if courts considered privacy policies to be contracts in the legal sense. Some scholars, including the authors of the Draft Restatement of the Law, Consumer Contracts, have argued that privacy policies usually constitute 'consumer contracts' (Draft Restatement, §1, Comment 9, Bar-Gill *et al.* (2017)). This classification as consumer contracts can impose restrictions on the ease with which services can change their privacy policies in a legally valid way. In particular, modifications of privacy policies may require 'assent'[24] According to the Draft Restatement, absent a contractually stipulated procedure governing modifications to a consumer contract, businesses are required to provide consumers with notice about a change as well as "a reasonable opportunity to reject the proposed modified term and continue the contractual relationship under the existing term" (Draft Restatement, §3). This requirement *de facto*

[24]The Draft Restatement on Consumer Contracts and Bar-Gill *et al.* (2017) provide a further discussion of the differences between consent and assent. In this paper we refer loosely to both standards as 'consent' primarily because all notions of consent in this paper refer to the procedure of agreeing to the terms of a consumer contract.

implies that privacy policies that lack a clause allowing for unilateral modifications by a business cannot be changed without the consumer's consent. The Draft Restatement has yet to be adopted. At the same time, some scholars dispute the Draft Restatement's assertion that existing case law supports the notion that privacy policies are contracts (Klass (2019); Levitin *et al.* (2019)).

2.3.2 Data

In the analysis, we use a dataset consisting of the texts of the privacy policies of 238 websites. For each website, we obtain one observation from April 2018 and one observation from April 2020.[25] To create this dataset, we begin with all websites that appeared in the Top 500 ranking in Alexa's Top Site service in late 2017.[26] We exclude websites operated by online services that are not headquartered in the United States.[27] The resulting dataset consists of 325 websites and their privacy policies. Notably, as these websites are operated by services located in the United States, they were not legally obliged to treat U.S. consumers in line with the requirements stipulated in the GDPR (GDPR, Art. 3(1), Frankenreiter (2021)).

We construct our treatment and control groups from the remaining 325 websites. Our treatment group consists of the U.S. versions of privacy policies of websites that either cater to E.U. consumers alongside consumers in other jurisdictions or offer one or more separate 'E.U. versions' of their websites.[28] Importantly, because of their substantial exposure to E.U. consumers, these services have to comply with the GDPR insofar as they serve their customers in the European Union (GDPR, Art. 3(2)). Yet we focus on privacy policies used in interactions between these services and consumers in the United States, where the GDPR

[25] Frankenreiter (2021) and Frankenreiter and Hermstruewer (2021) provide more details on how the privacy policies were obtained.

[26] `https://www.alexa.com/topsites`.

[27] We further restrict the sample to websites that have policies in English and do not use privacy policy as another website in the sample.

[28] One example in the former category is facebook.com, which is available in many languages primarily spoken in the European Union. The latter category includes Google, which makes its services available to E.U. residents mainly through country-specific versions of its website such as google.de.

does not apply. The treatment sample consists of 130 websites, corresponding to 260 privacy policies.

Our control group contains of the privacy policies of websites that we consider not to have been 'exposed' to the GDPR, meaning that they do not have to comply with the GDPR in any of their interactions with consumers. The reason for the lack of exposure to the GDPR is that these websites are not 'targeting' consumers in the E.U. in the way described above.[29] We exclude from our control sample all websites of services in industries that are covered by sector-specific privacy regulations in the U.S.[30] We furthermore exclude all websites that, although they do not explicitly target E.U. consumers, receive more than 10% of their traffic from the E.U., or for which the ratio of E.U. versus U.S. traffic is above 0.11.[31] We are left with a sample of 108 'control' websites and 216 privacy policies. The final dataset therefore consists of 476 policies from 238 websites. In addition to the texts of the privacy policies, we further obtain a range of background variables for all websites.[32]

Our main analysis focuses on a binary classification of E.U. exposure. In practice, however, exposure to E.U. regulation could be a matter of degree. A firm that caters primarily to E.U. subjects may have greater exposure to the GDPR than firms that caters primarily to U.S. consumers. In our Appendix, we therefore replicate the analysis presented here using a measure of E.U. exposure that is not binary, but measures the degree of E.U. exposure. For this, we obtain information on the overall percentage of traffic to a website comes from E.U. countries.[33]

We coded the policies in this dataset according to a coding scheme developed specifically for this project. The coding scheme captures various features of a website's privacy policy,

[29]Examples of websites in this category include large retailers without operations in the E.U. such as Lowes and Bed, Bath & Beyond.

[30]The industries we exclude are financial services, academia, government and medical.

[31]To make this determination, we use traffic data obtained from Alexa Web services in the summer of 2020. Frankenreiter (2021) describes this data in more detail. We exclude websites from our sample for which such traffic data is not available.

[32]This includes information such as the website category and website entry data.

[33]The estimates of the percentage of website traffic from E.U. countries are from October 2020.

including (1) the types of data that a service is allowed to gather, (2) the different ways in which a service is allowed to use data under the policy, and (3) rules related to policy changes. With regard to (1) and (2), the coding scheme contains lists of different types of information, and different information uses whose legal status under the policy we ascertain through our coding. Examples of items in the list for (1) are personally identifiable information and geolocation data. The list for (2) included data sharing with third parties for advertisement and marketing and data sharing for other activities external to the website. For each item, we asked our coder to ascertain whether the collection of this type of data or the information use was (i) explicitly allowed under the policy, (ii) implicitly permitted under the policy (because, for example, the policy contained open-ended language that did not restrict the practice), or (iii) not allowed under the policy. With regard to (3), the coding scheme requires coders to identify whether the rules fell into one of the following categories: (i) No mention of consent, (ii) policy stipulates that continued use of the website is treated as consent, (iii) users need to actively consent to any changes. The coding was done in the summer of 2021 by a research assistant from Columbia Law School.

2.3.3 Results

We begin by reporting how U.S. privacy policies change after the GDPR came into effect, comparing GDPR exposed websites to websites not exposed to the GDPR. We then consider whether policies of GDPR exposed websites differ from non-exposed websites with respect to the type of information they collect and the ways they use this data. It is important to consider this second dimension of data collection and usage because to some extent broad data rights and consent requirements can be viewed as substitutes. A policy that sets down very broad data rights may not require material change in the future making strict consent requirements less consequential. We therefore consider differences in data rights articulated in the privacy policies before and after the GDPR comes into effect.

Consent

As described above, we distinguish between three types of consent provisions in policies. Most policies in our sample contain no mention of whether consumers must consent to changes to the privacy policy ('no mention').[34] For the policies that do require consent, they either stipulate that consumers consent to the updated privacy policy when they continue to use the website ('continued use as consent')[35] or they require consumers to actively consent to the new policy ('active consent')[36].

Table 2.1 shows the consent provisions in our sample, separately for 2018 and for 2020. Over half of all policies in the sample do not mention any consent requirement for changes to the policy policy (56.4% in 2018 and 59.4% in 2020). The vast majority of policies that do set down rules for consenting to changes in policies state that continued use of the websites constitutes consent to the policy changes ('continued use as consent'), rather than requiring that consumers actively consent to changes.

---

[34]For example, in the 2018 zappos.com privacy policy there is mention of the possibility that the privacy policy can change but no provisions related to consumer consent: 'If we change or update this Privacy Policy, we will post changes and updates on the Site so that you will always be aware of what information we collect, use and disclose. We encourage you to review this Privacy Policy from time to time so you will know if the Privacy Policy has been changed or updated. If you have any questions about the Privacy Policy, please contact us at 1-800-943-1633 or cs@zappos.com.'

[35]For example, in Twitter's 2018 it says with respect to changes in the policy: 'We may revise this Privacy Policy from time to time. The most current version of the policy will govern our use of your information and will always be at https://twitter.com/privacy. If we make a change to this policy that, in our sole discretion, is material, we will notify you via an @Twitter update or email to the email address associated with your account. By continuing to access or use the Services after those changes become effective, you agree to be bound by the revised Privacy Policy.'

[36]For example, in Verizon's 2018 policy it says with respect to changes in the policy: 'We reserve the right to make changes to this privacy policy, so please check back periodically for changes. You will be able to see that changes have been made by checking to see the effective date posted at the end of the policy. If Verizon elects to use or disclose information that identifies you as an individual in a manner that is materially different from that stated in our policy at the time we collected that information from you, we will give you a choice regarding such use or disclosure by appropriate means, which may include use of an opt out mechanism.' Note that this provision only requires active consent in a narrow set of circumstances. We nonetheless code this as 'active consent' because it does provide for circumstances under which active consent it required.

Table 2.1: *Consent provisions*

	2018 consent			
	Active consent	Continued use as consent	No mention of consent	Total
	(1)	(2)	(3)	
No GDPR Exposure	1 (1%)	44 (41%)	63 (58%)	108
GDPR Exposure	1 (1%)	57 (44%)	72 (55%)	130
Total	2 (1%)	101 (42%)	135 (57%)	238
	2020 consent			
	Active consent	Continued use as consent	No mention of consent	Total
	(1)	(2)	(3)	
No GDPR Exposure	2 (2%)	50 (46%)	56 (52%)	108
GDPR Exposure	0	44 (34%)	86 (66%)	130
Total	2 (1%)	94 (39%)	142 (60%)	238

Changes in consent stringency

We are primarily interested in whether privacy policies are more or less stringent in their 2020 policy, following the GDPR coming into effect, relative to their 2018 policy. Most policies are unchanged with respect to their consent provision,[37] with 83.7% of websites with no GDPR exposure remaining with the same consent provision type in their 2018 and 2020 policies and 78.5% of websites with GDPR exposure remaining with the same consent provision type in their 2018 and 2020 policies.

Defining stringency.– We define policies as **less stringent** if in 2018 the policy required 'active consent' and in 2020 the policy either required 'continued use as consent' or there is 'no mention' of consent in the policy. A policy also becomes less stringent with respect to its consent requirement if in 2018 it required 'continued use as consent' and in 2020 it no longer mentions consent.[38] Conversely, a policy is **more stringent** with respect to its consent provision if the 2018 policy did not mention consent and the 2020 requires 'continued use as consent' or 'active consent'. The policy is also more stringent if in 2018 it required 'continued

[37]This does not necessarily mean that the exact language in the policy remained the same from 2018 to 2020 but that in essence the type of consent required has not changed even if the language has changed.

[38]As noted above, FTC jurisprudence suggests that that retroactive changes to policies as requiring informed consent it remains unclear what change would be considered retroactive. Moreover, websites themselves do not distinguish between prospective and retroactive changes to policies.

use as consent' and in 2020 required 'active consent.' It is important to note that although at first 'continued use as consent' could be viewed as close to having not consent requirement at all, in practice 'continued use as consent' could create significant barriers for websites in obtaining containing. For example, if a website uses data of a user who is no longer active on the website.

Our definitions of consent stringency rely on our understanding that firms that drop any mention of consent seek to lower their consent requirement, for example, by allowing the mere posting of a new policy to be sufficient for the new policy to apply. If privacy policies are considered contracts,[39] it is possible that a privacy policy that does not lay down a mechanism for policy changes would actually require active consent for policy changes, the default for consent for privacy policies. This could mean that from a legal perspective, dropping any mention of consent would create a **more stringent** consent requirement.

Despite ongoing debates on whether privacy policies are considered contracts, we adopt the interpretation that websites that drop any mention of consent intend to adopt a **less stringent** consent requirement. First, it continues to be disputed whether privacy policies are indeed contracts and in any event there is little case law on the exact implications of policy silence on required consent for policy changes. Second, we are primarily interested in website's subjective beliefs. Based on conversations with industry practitioners, we consider it unlikely that websites believe that by omitting mention of consent they are opting in to the most stringent consent requirement.

Consent stringency results.– Figure 2.1 shows the percentage of privacy policies that became **less stringent**, **more stringent**, or **unchanged** in their 2020 policy, reporting percentages separately for websites that are exposed to the GDPR and websites that are not exposed to the GDPR. Figure 2.1 shows that websites that are exposed to the GDPR are more likely to make their policies **less stringent** with respect to consent requirements while websites that are not exposed to the GDPR are more likely to make their consent requirements **more**

[39]See discussion in Section 2.3.1. Classifying privacy policies as contracts may also have implications for whether explicit consent provisions are enforceable (Bar-Gill and Davis (2010)).

stringent. While most policies do not make an essential change to their consent provisions, of the websites that do change their policy's consent provision there is a clear difference between GDPR exposed websites, who are more likely to choose to have less stringent policies in 2020 and websites with no GDPR exposure who are more likely to make their policies more stringent. Table 2.2 reports these results along with their statistical significance. In out Appendix we report the results using a logistical regression with the percentage of E.U. users' as the independent variable.

Figure 2.1: *Provision changes*

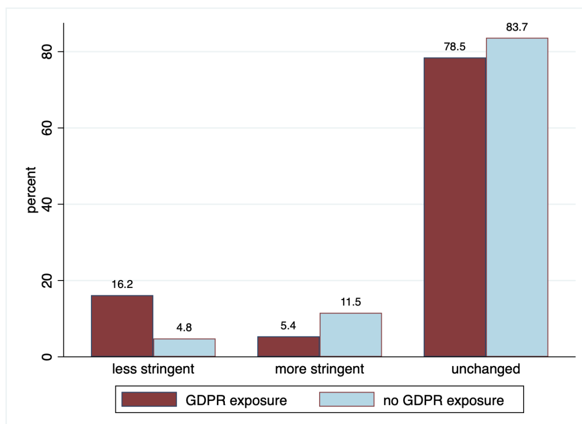


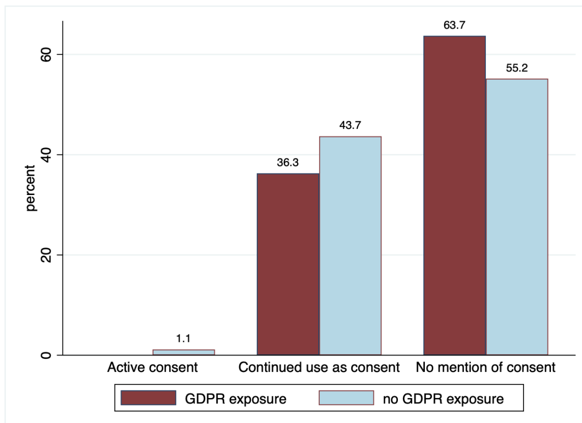
Table 2.2: *Consent provision changes*

	Consent Change			
	Less Stringent	More Stringent	Unchanged	Total
	(1)	(2)	(3)	
GDPR Exposure	21 (16.15%)	7 (5.38%)	102 (78.46%)	130
	14.4	10.6	105.0	
No GDPR Exposure	5 (4.81%)	12 (11.54%)	87 (83.65%)	104
	11.6	8.4	84.0	
Total	26 (11.11%)	19 (8.12%)	189 (80.77%)	234

Pearson Chi-square = 9.5818, P-value = 0.008, Fisher's exact = 0.007

For GDPR-exposed websites, most policies that do not change their consent provisions offer the lowest level of consent stringency. Figure 2.2 shows the consent provisions for policies that did not change from 2018 to 2020. The graph shows that 63.7% of unchanged GDPR exposed policies do no mention consent at all. This is important because 'no mention' is a lower bound for consent provisions, and therefore the results could understate a website's preference for less stringent consent provisions post GDPR, as many policies already had the lowest consent stringency possible.

Figure 2.2: *Provisions for policies that are unchanged*



A key challenge to any comparison between GDPR exposed websites and non-exposed websites is that whether a website is exposed to the GDPR is not random. The treatment and control groups vary with respect to their industry or category type. Moreover, the treatment and control group are not identical with respect to their user numbers even though both control and treatment groups contain high traffic websites.

We build on our analysis of changes in consent provisions of the full sample by considering the differences in a matched sample. To create this matched sample we identify all treatment websites that have an Alexa category of which there is at least one control website with the same category. We end up with a sample of 96 treatment firms. When there is more than one website in the control sample with the same category as a treatment website we select a match, with replacement, based on similarity in user numbers. Table 2.3 shows that also within this matched sample GDPR exposed websites are more likely to adopt a less stringent consent requirement for policy changes.

Table 2.3: *Consent provision changes: Matched sample*

	Consent Change			
	Less Stringent	More Stringent	Unchanged	Total
	(1)	(2)	(3)	
GDPR Exposure	16 (16.7%)	5 (5.2%)	75 (78.1%)	96
	11.0	4.0	81.0	
No GDPR Exposure	6 (6.3%)	3 (3.1%)	87 (90.6%)	96
	11.0	4.0	81.0	
Total	22 (11.5%)	8 (4.2%)	162 (84.4%)	192

Pearson Chi-square = 5.9343, P-value = 0.051

Data collection and usage

On the dimension of the scope of permissible data gathering, we see a move towards policies that explicitly allow for the gathering of certain sensitive types of data. This is true for both websites of services that were exposed to the GDPR and for websites of services that were not. As an example, Table 2.4 displays the results for whether privacy policies allow for the gathering of geolocation data, separated by the year in which the policies were gathered. Geolocation data is a particularly sensitive type of data because services (especially those that users access frequently using their mobile devices) can compile comprehensive movement profiles of their users.

We can see that only 1 out of 238 privacy policies adopted a provision restricting the collection of geolocation data prior to when the GDPR entered into force. In 2020, only 2 out of 238 services had provisions that outlawed the use of such data. At the same time, a substantial number of services went from implicitly allowing for the geolocation data in 2018 to explicitly allowing for the collection of geolocation data in 2020. In 2018, most services already allowed for the collection of such data, but 24% of websites had privacy policies in place that included broad, open-ended clauses that arguably covered the collection of many types of data, including geolocation data. By 2020, the number of services without a clause explicitly allowing for such data collection had dropped to less than half.

Importantly, this trend can be observed both for websites exposed to the GDPR and

Table 2.4: *Geolocation data collection provisions*

	2018 geodata			
	Not allowed	Implicitly allowed	Explicitly allowed	Total
	(1)	(2)	(3)	
No GDPR Exposure	1 (1%)	24 (22%)	83 (77%)	108
GDPR Exposure	0	33 (25%)	97 (75%)	130
Total	1 (0%)	57 (24%)	180 (76%)	238

	2020 geodata			
	Not allowed	Implicitly allowed	Explicitly allowed	Total
	(1)	(2)	(3)	
No GDPR Exposure	2 (2%)	12 (11%)	94 (87%)	108
GDPR Exposure	0	12 (9%)	118 (91%)	130
Total	2 (1%)	24 (10%)	212 (89%)	238

those that were not. A larger share of exposed services adopted explicit provisions allowing for the collection of geolocation data. However, the difference is slight (and not statistically significant).

Similar trends can be observed for privacy policy provisions addressing the scope of permissible data uses. To illustrate these changes, Table 2.5 displays results for whether privacy policies allow for the sharing of data with third-party advertisement services. Similar to the provisions regarding geolocation data described above, the number of privacy policies with provisions explicitly allowing for this sharing of data grew over the two years covered in our analysis. We again see this trend for privacy policies of services exposed to the GDPR and those not. Similar trends can be observed for all other types of information and data uses for which information is available in our dataset.

2.3.4 Discussion

Our results show different trends in consent provisions for websites exposed to the GDPR and websites that are not exposed to the GDPR. Websites that are exposed to the GDPR are more likely to adjust their policies post-GDPR to contain a less stringent consent provision.

Table 2.5: *Third-party advertisers provisions*

	2018 third-party advertising			
	Not allowed	Implicitly allowed	Explicitly allowed	Total
	(1)	(2)	(3)	
No GDPR Exposure	2 (2%)	19 (18%)	87 (81%)	108
GDPR Exposure	5 (5%)	18 (14%)	107 (82%)	130
Total	7 (3%)	37 (16%)	194 (82%)	238
	2020 third-party advertising			
	Not allowed	Implicitly allowed	Explicitly allowed	Total
	(1)	(2)	(3)	
No GDPR Exposure	0	10 (9%)	98 (91%)	108
GDPR Exposure	3 (3%)	6 (5%)	121 (93%)	130
Total	3 (1%)	16 (7%)	219 (92%)	238

Conversely, websites that are not exposed to the GDPR are more likely to adjust their policies post-GDPR to contain more stringent consent provisions. This is in striking contrast to the dimensions related to data collection collection in usage in which both GDPR exposed and non-exposed websites exhibit a similar trend of moving from implicit data rights to explicit data rights.

Our interpretation of what constitutes a more or less stringent consent requirement is not obvious. Most websites that change their consent provisions either drop any mention of consent from their privacy policies (in most cases because they drop their 'continued use as consent' provision) or by adding a 'continued use as consent' provision (when previously the policy did not mention consent at all). Viewing privacy policies as contracts, a position taken by the Draft Restatement on Consumer Contracts and some in the literature (Bar-Gill *et al.* (2017)), suggests that dropping a mechanism for policy changes means consumers must then **actively consent** to a policy change. A policy that drops any mention of consent is possibly opting into a stricter consent regime of active consent. Such a result would lead to the opposite conclusion to what we state above – firms exposed to the GDPR move closer to the GDPR in their U.S. policies by dropping mention of consent.

We believe it is unlikely that U.S. websites that drop any mention of consent intend to create a more stringent consent requirement. Firstly, it continues to be a topic of contention whether privacy policies should be considered contracts, with some scholars arguing that case law does not indicate a clear trend of recognizing policies as contracts (Klass (2019)). Moreover, there are no cases that directly consider this question of policy updating and FTC guidance suggests that only retroactive changes require active consent. Lastly, even if a court would determine that a silent privacy policy requires active consent for policy changes, we are primarily interested in firm responses to the GDPR which rely on the beliefs of firms. Based on our conversations with industry practitioners, we consider it unlikely that websites that drop any mention of consent intend to increase consent stringency.

We interpret these results as suggestive that websites through their exposure to the GDPR learn about the impact of the GDPR's consent requirement. These websites may learn that requiring 'active consent' for privacy policy changes could be particularly costly if a critical number of consumers do not consent to the changes, preventing websites from using and sharing consumer data in ways not originally stipulated in the policy. Even a more lenient consent requirement of 'continued use as consent' could be costly if a website has data of consumers that are no longer active on the website. Websites that experience these costs may then choose to avoid consent requirements in policies that are not subject to the GDPR, and therefore adjust their U.S. policies to have less stringent consent provisions.

Our findings are consistent with prior research, which suggests that the GDPR's consent requirement has had implications for consumer behavior and website traffic. Schmitt *et al.* (2021) show that in the 18 months after the GDPR came into effect, the average number of unique website visitors declined for the $6,286$ most popular websites in E.U. countries. This result suggests that privacy features, such as requesting consumers explicit consent for data collection, effects consumer behavior. Similarly, Aridor *et al.* (2020) show that the ability to opt-out of cookie tracking lead to a reduction of cookies collected by websites. These results suggest that the GDPR's consent requirement may have real implications for website's ability to obtain consent.

A learning story is not the only way to understand the move of GDPR exposed firms to less stringent consent requirements. For example, the need to comply with the GDPR may have prompted GDPR exposed firms to revise and update their privacy policies, where non-exposed firms may opt to continue with their existing policy. This explanation is consistent with the slightly higher number of policies with consent provisions changes for GDPR exposed firms (21.53% for GDPR exposed firms versus 16.35% for non-GDPR exposed firms) although this would not explain why firms not exposed to the GDPR opt into stricter consent regimes.

These results also reflect a specific context in which the GDPR has not raised privacy policy standards for policies not directly subject to the GDPR. The theory of unilateral regulatory globalization, commonly known as the 'Brussels Effect' (Bradford (2020)), refers to a situation in which firms comply with E.U. law even when not directly subject to it. If it is costly for businesses to treat consumers in different jurisdictions differently, this could lead to firms with consumers in the E.U. to treat everyone in accordance with GDPR standards. However, in this context we see that websites with GDPR exposure choose to be less stringent with respect to consent.

Chapter 3

Big Data and Discrimination

Abstract

The ability to distinguish between people in setting the price of credit is often constrained by legal rules that aim to prevent discrimination. These legal requirements have developed focusing on human decision-making contexts, and so their effectiveness is challenged as pricing increasingly relies on intelligent algorithms that extract information from big data. In this Essay, we bring together existing legal requirements with the structure of machine-learning decision-making in order to identify tensions between old law and new methods and lay the ground for legal solutions. We argue that, while automated pricing rules provide increased transparency, their complexity also limits the application of existing law. Using a simulation exercise based on real-world mortgage data to illustrate our arguments, we note that restricting the characteristics that the algorithm is allowed to use can have a limited effect on disparity and can in fact increase pricing gaps. Furthermore, we argue that there are limits to interpreting the pricing rules set by machine learning that hinders the application of existing discrimination laws. We end by discussing a framework for testing discrimination that evaluates algorithmic pricing rules in a controlled environment. Unlike the human decision-making context, this framework allows for ex ante testing of price rules, facilitating comparisons between lenders.

3.1 Introduction

For many financial products, such as loans and insurance policies, companies distinguish between people based on their different risks and returns. However, the ability to distinguish between people by trying to predict future behavior or profitability of a contract is often restrained by legal rules that aim to prevent certain types of discrimination. For example, the Equal Credit Opportunity Act[1] (ECOA) forbids race, religion, age, and other factors from being considered in setting credit terms,[2] and the Fair Housing Act[3] (FHA) prohibits discrimination in financing of real estate based on race, color, national origin, religion, sex, familial status, or disability.[4] Many of these rules were developed to challenge human discretion in setting prices and provide little guidance in a world in which firms set credit terms based on sophisticated statistical methods and a large number of factors. This rise of artificial intelligence and big data raises the questions of when and how existing law can be applied to this novel setting, and when it must be adapted to remain effective.

In this Essay, we bridge the gap between old law and new methods by proposing a framework that brings together existing legal requirements with the structure of algorithmic decision-making in order to identify tensions and lay the ground for legal solutions. Focusing on the example of credit pricing, we confront steps in the genesis of an automated pricing rule with their regulatory opportunities and challenges.

Based on our framework, we argue that legal doctrine is ill prepared to face the challenges posed by algorithmic decision-making in a big data world. While automated pricing rules promise increased transparency, this opportunity is often confounded. Unlike human decision-making, the exclusion of data from consideration can be guaranteed in the algo-

[1] Pub L No 94-239, 90 Stat 251 (1976), codified as amended at 15 USC § 1691 et seq.

[2] 15 USC § 1691(a)(1)–(3).

[3] Pub L No 90-284, 82 Stat 81 (1968), codified as amended at 42 USC § 3601 et seq.

[4] 42 USC § 3605(a). These laws do not exhaust the legal framework governing discrimination in credit pricing. Beyond other federal laws that also relate to credit discrimination, such as the Community Reinvestment Act, Pub L No 95-128, 91 Stat 1111 (1977), codified at 12 USC § 2901 et seq, there are many state and local laws with discrimination provisions, such as fair housing laws.

rithmic context. However, forbidding inputs alone does not assure equal pricing and can even increase pricing disparities between protected groups. Moreover, the complexity of machine-learning pricing limits the ability to scrutinize the process that led to a pricing rule, frustrating legal efforts to examine the conduct that led to disparity. On the other hand, the reproducibility of automated prices creates new possibilities for more meaningful analysis of pricing outcomes. Building on this opportunity, we provide a framework for regulators to test decision rules ex ante in a way that provides meaningful comparisons between lenders.

To consider the challenges to applying discrimination law to a context in which credit pricing decisions are fully automated, we consider both the legal doctrine of "disparate treatment," dealing with cases in which a forbidden characteristic is considered directly in a pricing decision, and "disparate impact," when a facially neutral conduct has a discriminatory effect.[5] While in general the availability of a disparate impact claim depends on the legal basis of the discrimination claim, in the context of credit pricing, the law permits the use of disparate impact as a basis of a discrimination claim both under the FHA and the ECOA.[6] A comprehensive discussion of these two doctrines and their application to credit pricing is beyond the scope of this Essay, particularly because there are several aspects of these doctrines on which there is widespread disagreement.[7] We therefore abstract away from some of the details of the doctrines and focus on the building blocks that create a discrimination claim. Developing doctrine that is appropriate for this context ultimately

[5]For an overview of these two legal doctrines and their relation to theories of discrimination, see John J. Donohue, *Antidiscrimination Law*, in A. Mitchell Polinsky and Steven Shavell, 2 *Handbook of Law and Economics* 1387, 1392–95 (Elsevier 2007).

[6]The Supreme Court recently affirmed that disparate impact claims could be made under the FHA in *Texas Department of Housing and Community Affairs v Inclusive Communities Project, Inc*, 135 S Ct 2507, 2518 (2015), confirming the position of eleven appellate courts and various federal agencies, including the Department of Housing and Urban Development (HUD), which is primarily responsible for enforcing the FHA. See also generally Robert G. Schwemm, *Fair Housing Litigation after* Inclusive Communities*: What's New and What's Not*, 115 Colum L Rev Sidebar 106 (2015). Although there is not an equivalent Supreme Court case with respect to the ECOA, the Consumer Financial Protection Bureau and courts have found that the statute allows for a claim of disparate impact. See, for example, *Ramirez v GreenPoint Mortgage Funding, Inc*, 633 F Supp 2d 922, 926–27 (ND Cal 2008).

[7]For further discussion of the discrimination doctrines under ECOA and FHA, see Michael Aleo and Pablo Svirsky, *Foreclosure Fallout: The Banking Industry's Attack on Disparate Impact Race Discrimination Claims under the Fair Housing Act and the Equal Credit Opportunity Act*, 18 BU Pub Int L J 1, 22–38 (2008); Alex Gano, Comment, *Disparate Impact and Mortgage Lending: A Beginner's Guide*, 88 U Colo L Rev 1109, 1128–33 (2017).

requires a return to the fundamental justifications and motivations behind discrimination law.

Specifically, we consider three approaches to discrimination.[8] The first approach is to focus on the "inputs" of the decision, stemming from the view that discrimination law is primarily concerned with formal or intentional discrimination.[9] The second approach scrutinizes the decision-making process, policy, or conduct that then led to disparity. The third approach focuses on the disparity of the "outcome."[10] We consider the options facing a social planner to achieve different policy ends and discuss how algorithmic decision-making challenges each of these options, without adopting a particular notion of discrimination.

Existing legal doctrine provides little guidance on algorithmic decision-making because the typical discrimination case focuses on the human component of the decision, which often remains opaque. Consider a series of cases from around 2008 that challenged mortgage pricing practices. In these cases, plaintiffs argued that black and Hispanic borrowers ended up paying higher interest rates and fees after controlling for the "par rate" set by the mortgage originator. The claim was that the discretion given to the mortgage originator's employees and brokers in setting the final terms of the loans above the "par rate," and the incentives to do so, caused the discriminatory pricing.[11] These types of assertions were

[8] We find it necessary to divide approaches to discrimination by their goal and focus because the doctrines of disparate treatment and disparate impact can be consistent with more than one approach depending on the exact interpretation and implementation of the doctrine. Moreover, legal doctrines often require more than one approach to demonstrate a case for disparate impact or disparate treatment, such as in the three-part burden-shifting framework for establishing an FHA disparate impact case as formulated by HUD. See Implementation of the Fair Housing Act's Discriminatory Effects Standard, 78 Fed Reg 11459, 11460–63 (2013), amending 24 CFR § 100.500.

[9] This basic articulation is also used in Richard Primus, *The Future of Disparate Impact*, 108 Mich L Rev 1341, 1342 (2010). For a discussion on the different notions of intention, see Aziz Z. Huq, *What Is Discriminatory Intent?*, 103 Cornell L Rev 1212, 1240–65 (arguing that judicial theory of "intention" is inconsistent).

[10] We do not argue directly for any of these three approaches; rather, we point to the opportunities and challenges that machine learning credit pricing creates for each approach.

[11] Most of these cases are disparate impact cases, although some of them are more ambiguous as to the exact grounds for the discrimination case and may be read as disparate treatment cases.

made in the context of individual claims,[12] class actions,[13] and regulatory action.[14] What is most striking is that these cases do not directly scrutinize the broker decisions, treating them as a "black box," but focus instead on the mortgage originator's discretion policy.[15] Had the courts been able to analyze the discriminatory decisions directly, we would have had a greater understanding of the precise conduct that was problematic. As a result of the scope and range of the legal doctrine, which are important for the automated pricing context, discrimination cases that involve opaque human decisions do not allow us to develop the exact perimeters of the doctrine.[16]

When algorithms make decisions, opaque human behavior is replaced by a set of rules constructed from data. Specifically, we consider prices that are set based on prediction of mortgage default. An algorithm takes as an input a training data set with past defaults and then outputs a function that relates consumer characteristics, such as their income and credit score, to the probability of default. Advances in statistics and computer science have

[12]See, for example, *Martinez v Freedom Mortgage Team, Inc*, 527 F Supp 2d 827, 833–35 (ND Ill 2007).

[13]See, for example, *Ramirez*, 633 F Supp 2d at 924–25; *Miller v Countrywide Bank, National Association*, 571 F Supp 2d 251, 253–55 (D Mass 2008).

[14]For a discussion of a series of complaints by the Justice Department against mortgage brokers that were settled, see Ian Ayres, Gary Klein, and Jeffrey West, *The Rise and (Potential) Fall of Disparate Impact Lending Litigation*, in Lee Anne Fennell and Benjamin J. Keys, eds, *Evidence and Innovation in Housing Law and Policy* 231, 240–46 (Cambridge 2017) (discussing cases against Countrywide (2011), Wells Fargo (2012), and Sage Bank (2015), all involving a claim that discretion to brokers resulted in discrimination).

[15]The use of a discretion policy as the conduct that caused the discriminatory effect has been applied by the CFPB to ECOA cases. See, for example, Consent Order, *In the Matter of American Honda Finance Corporation*, File No 2015-CFPB-0014, *5–9 (July 14, 2015) (available on Westlaw at 2015 WL 5209146). This practice has also been applied in other areas, such as employment discrimination cases. For example, the seminal employment discrimination case *Watson v Fort Worth Bank & Trust*, 487 US 977, 982–85 (1988), dealt with a disparate impact claim arising from discretionary and subjective promotion policies. The future of these types of class action cases is uncertain given *Wal-Mart Stores, Inc v Dukes*, 564 US 338, 352–57 (2011) (holding that, in a suit alleging discrimination in Wal-Mart's employment promotion policies, class certification was improper because an employer's discretionary decision-making is a "presumptively reasonable way of doing business" and "merely showing that [a company's] policy of discretion has produced an overall sex-based disparity does not suffice" to establish commonality across the class.).

[16]It is important to note that this opaqueness is not only evidentiary, meaning the difficulty in proving someone's motivation and intentions in court. It is also a result of human decision-making often being opaque to the decisionmakers themselves. There are decades of research showing that people have difficulty recovering the basis for their decisions, particularly when they involve race. See, for example, Cheryl Staats, et al, *State of the Science: Implicit Bias Review 2015* *4–6 (Kirwan Institute for the Study of Race and Ethnicity, 2015), archived at http://perma.cc/4AJ6-4P4C.

produced powerful algorithms that excel at this prediction task, especially when individual characteristics are rich and data sets are large. These machine-learning algorithms search through large classes of complex rules to find a rule that works well at predicting the default of new consumers using past data. Because we consider the translation of the default prediction into a price as a simple transformation of the algorithm's prediction, we refer to the prediction and its translation into a pricing rule jointly as the "decision rule."[17]

We connect machine learning, decision rules, and current law by considering the three stages of a pricing decision, which we demonstrate in a simulation exercise. The data we use is based on real data on mortgage applicants from the Boston HMDA data set,[18] and we impute default probabilities from a combination of loan approvals and calibrate them to overall default rates.[19] The simulated data allows us to demonstrate several of our conceptual arguments and the methodological issues we discuss. However, given that crucial parts of the data are simulated, the graphs and figures in this Essay should not be interpreted as reflecting real-world observations but rather methodological challenges and opportunities that arise in the context of algorithmic decision making.[20]

The remainder of this Essay discusses each of the three steps of a pricing decision by underlining both the challenges and the opportunities presented by applying machine-learning pricing to current legal rules. First, we consider the data input stage of the pricing decision

[17]We assume that prices are directly obtained from predictions, and our focus on predicted default probabilities is therefore without loss of generality. Other authors instead consider a separate step that links predictions to decisions. See, for example, Sam Corbett-Davies, et al, *Algorithmic Decision Making and the Cost of Fairness*, Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining *2–3 (ArXiv, 2017), archived at http://perma.cc/9G5S-JDT8. In order to apply our framework to such a setup, we would directly consider the resulting pricing rule.

[18]Mortgage originators are required to disclose mortgage application information under the Home Mortgage Disclosure Act (HMDA), Pub L No 94–200, 89 Stat 1124 (1975), codified at 12 USC § 2801 et seq, including applicant race. The Boston HMDA data set combines data from mortgage applications made in 1990 in the Boston area with a follow-up survey collected by the Federal Reserve Bank of Boston. See Alicia H. Munnell, et al, *Mortgage Lending in Boston: Interpreting HDMA Data*, 86 Am Econ Rev 25, 30–32 (1996). Further information on the data set can be found in an online appendix.

[19]The HMDA data set includes only applicant status, so we need to simulate default rates to engage in a default prediction exercise. The calibration of overall default rates is based on Andreas Fuster, et al, *Predictably Unequal? The Effects of Machine Learning on Credit Markets* (2018), archived at http://perma.cc/LYY5-SAG2.

[20]The Appendix contains more details on how this data was constructed.

and argue that excluding forbidden characteristics has limited effect and satisfies only a narrow understanding of antidiscrimination law.[21] One fundamental aspect of antidiscrimination laws is the prohibition on conditioning a decision on the protected characteristics, which can formally be achieved in automated decision-making. However, the exclusion of the forbidden input alone may be insufficient when there are other characteristics that are correlated with the forbidden input—an issue that is exacerbated in the context of big data.[22] In addition, we highlight the ways in which restricting a broader range of data inputs may have unintended consequences, such as increasing price disparity.[23]

Second, we connect the process of constructing a pricing rule to the legal analysis of conduct and highlight which legal requirements can be tested from the algorithm.[24] This stage of the firm's pricing decision is often considered the firm's "conduct," which can be scrutinized for identifying the particular policy that led to the disparity. Unlike the context of human decision-making, in which conduct is not fully observed, in algorithmic decision-making we are able to observe the decision rule. We argue, however, that this transparency is limited to the types of issues that are interpretable in the algorithmic context. In particular, many machine-learning methods do not allow a general-purpose determination about which variables are important for the decision rule absent further clarification regarding what "importance" would mean in this legal context.

Third, we consider the statistical analysis of the resulting prices and argue that the observability of the decision rules expands the opportunities for controlled and preemptive testing of pricing practices.[25] The analysis of the outcome becomes attractive in the context of algorithmic decision-making given the limitations of an analysis of the input and decision process stage. Moreover, outcome analysis in this new context is not limited to actual prices

[21] See Section 3.3.

[22] See Section 3.3.1.

[23] See Section 3.3.2.

[24] See Section 3.4.

[25] See Section 3.5.

paid by consumers as we are able to observe the decision rule for future prices, allowing for forward-looking analysis of decision rules. This type of analysis is especially useful for regulators that enforce antidiscrimination law.

Our framework contributes to bridging the gap between the literature on algorithmic fairness and antidiscrimination law. Recent theoretical, computational, and empirical advances in computer science and statistics provide different notions of when an algorithm produces fair outcomes and how these different notions relate to one another.[26] However, many of these contributions focus solely on the statistical analysis of outcomes but neither explicitly consider other aspects of the algorithmic decision process nor relate the notions of fairness to legal definitions of discrimination.[27] By providing a framework that relates the analysis of algorithmic decision-making to legal doctrine, we highlight how results from this literature can inform future law through the tools it has developed for the statistical analysis of outcomes.

3.2 Setup for Illustration and Simulation

Throughout this Essay, we consider the legal and methodological challenges in analyzing algorithmic decision rules in a stylized setting that we illustrate with simulated data. In our example, a firm sets loan terms for new consumers based on observed defaults of

---

[26]See, for example, Jon Kleinberg, et al, *Algorithmic Fairness*, 108 AEA Papers and Proceedings 22, 22–23 (2018) (arguing that "across a wide range of estimation approaches, objective functions, and definitions of fairness, the strategy of blinding the algorithm to race inadvertently detracts from fairness.").

[27]There are some exceptions. See, for example, Michael Feldman, et al, *Certifying and Removing Disparate Impact, Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* *2–3 (ArXiv, 2015). Although the paper attempts to provide a legal framework on algorithmic fairness, its focus is on the Equal Employment Opportunity Commission's 80 percent rule and fairness as the ability to predict the protected class. The paper therefore does not capture the most significant aspects of antidiscrimination law. Prior literature has suggested that big data may pose challenges to antidiscrimination law, particularly for Title VII employment discrimination. See, for example, Solon Barocas and Andrew D. Selbst, *Big Data's Disparate Impact*, 104 Cal L Rev 671, 694–714 (2016) (focusing on several channels, primarily through biased human discretion in the data generating process, in which the data mining will reinforce bias). In contrast, our argument applies even when there is no human bias in past decisions. For a paper focused on the issues that big data cases pose for antidiscrimination law in the context of credit scores, see generally Mikella Hurley and Julius Adebayo, *Credit Scoring in the Era of Big Data*, 18 Yale J L & Tech 148 (2016) (focusing on the transparency issues created by big data that will limit people's ability to challenge their credit score).

past clients. Specifically, the company learns a prediction of loan default as a function of individual characteristics of the loan applicant from a training sample. It then applies this prediction function to new clients in a held-out data set. This setup would be consistent with the behavior of a firm that aims to price loans at their expected cost.

In order to analyze algorithmic credit pricing under different constraints, we simulate such training and holdout samples from a model that we have constructed from the Boston HMDA data set. While this simulated data includes race identifiers, our model assumes that race has no direct effect on default.[28] We calibrate overall default probabilities to actual default probabilities from the literature, but because all defaults in this specific model are simulated and not based on actual defaults, any figures and numerical examples in this Essay should not be seen as reflecting real-world observations. Rather, our simulation illustrates methodological challenges in applying legal doctrine to algorithmic decision-making.

In the remainder of this Essay, we highlight methodological challenges in analyzing algorithmic decision-making by considering two popular machine-learning algorithms, namely the random forest and the lasso. Both algorithms are well-suited to obtain predictions of default from a high-dimensional data set. Specifically, we train both algorithms on a training sample with 2,000 clients with approximately 50 variables each, many of which are categorical. We then analyze their prediction performance on a holdout data set with 2,000 new clients drawn from the same model. While these algorithms are specific, we discuss general properties of algorithmic decision-making in big data.

3.3 Data Inputs and Input-Focused Discrimination

One aspect of many antidiscrimination regimes is a restriction on inputs that can be used to price credit. Typically, this means that protected characteristics, such as race and gender, cannot be used in setting prices. Indeed, many antidiscrimination regimes include rules on the exclusion of data inputs as a form of discrimination prevention. For example, a

[28] Default rates may still differ between groups because individuals differ in other attributes across groups, but we assume in our model that the race identifier does not contribute variation beyond these other characteristics.

regulation implementing the ECOA provides: "Except as provided in the Act and this part, a creditor shall not take a prohibited basis into account in any system of evaluating the creditworthiness of applicants."[29] Moreover, the direct inclusion of a forbidden characteristic in the decision process could trigger the disparate treatment doctrine because the forbidden attribute could directly affect the decision.

Despite the centrality of input restriction to discrimination law, the enforcement of these rules is difficult when the forbidden attribute is observable to the decision maker.[30] When a decision maker, such as a job interviewer or mortgage broker, observes a person's race, for example, it is impossible to rule out that this characteristic played a role in the decision, whether consciously or subconsciously. The most common type of credit disparate impact case deals with situations in which there is a human decision maker,[31] meaning that it is impossible to prove that belonging to a protected group was not considered. As we discuss in the Introduction, in the series of mortgage lending cases in which mortgage brokers had discretion in setting the exact interest and fees of the loan, it is implicit that customers' races were known to the brokers who met face-to-face with the customers. Therefore, we cannot rule out that race was an input in the pricing outcome.

The perceived opportunity for algorithmic decision-making is that it allows for formal exclusion of protected characteristics, but we argue that it comes with important limitations. When defining and delineating the data that will be used to form a prediction, we can guarantee that certain variables or characteristics are excluded from the algorithmic decision. Despite this increased transparency that is afforded by the automation of pricing, we show that there are two main reasons that discrimination regimes should not focus on input restriction. First, we argue that, if price disparity matters, input restriction is insufficient. Second, the inclusion of the forbidden characteristic may in fact decrease

[29] 12 CFR § 1002.6(b)(1).

[30] There is a further issue that we do not discuss, which is the inherent tension between excluding certain characteristics from consideration on the one hand and the requirement that a rule not have disparate impact, which requires considering those characteristics. For further discussion of this tension, see generally Richard A. Primus, *Equal Protection and Disparate Impact: Round Three*, 117 Harv L Rev 493 (2003).

[31] See Aleo and Svirsky, 18 BU Pub Int L J at 33–35.

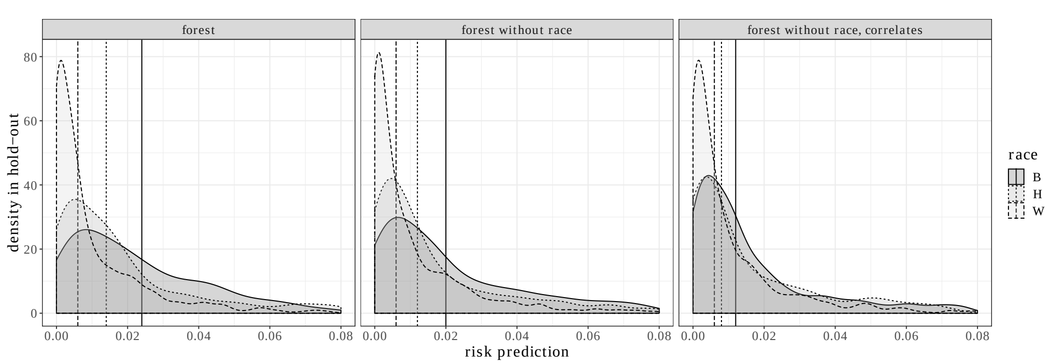
disparity, particularly when there is some measurement bias in the data.

3.3.1 Exclusion Is Limited

The formal exclusion of forbidden characteristics, such as race, would exclude any direct effect of race on the decision. This means that we would exclude any influence that race has on the outcome that is not due to its correlation with other factors. We would therefore hope that excluding race already reduces a possible disparity in risk predictions between race groups in algorithmic decision-making. However, when we exclude race from fitting the algorithm in our simulation exercise, we show below that there is little change in how risk predictions differ between protected groups.

To demonstrate that disparity can indeed persist despite the exclusion of input variables, consider the three graphs below (3.1) that represent the probability density function of the predicted default rates of the customers in a new sample not used to train the algorithm by race/ethnicity and using a random forest as a prediction algorithm. On the left, the distribution of predicted default rates was created using the decision rule that included the group identity as an input. We can see that the predicted default distributions are different for whites, blacks, and Hispanics. The median prediction for each group is represented by the vertical lines. The middle graph shows the distribution of predicted default rates when race is *excluded* as an input from the algorithm that produced the decision rule. Despite the exclusion of race, much of the difference among groups persists.

Figure 3.1: *Distribution of risk predictions across groups for different inputs*



Indeed, if there are other variables that are correlated with race, then predictions may strongly vary by race even when race is excluded, and disparities may persist. For example, if applicants of one group on average have lower education, and education is used in pricing, then using education in setting prices can imply different prices across groups. If many such variables come together, disparities may persist. In very high-dimensional data, and when complex, highly nonlinear prediction functions are used, this problem that one input variable can be reconstructed jointly from the other input variables becomes ubiquitous.

One way to respond to the indirect effect of protected characteristics is to expand the criteria for input restriction. For example, if an applicant's neighborhood is highly correlated with an applicant's race, we may want to restrict the use of one's neighborhood in pricing a loan. A major challenge of this approach is the required articulation of the conditions under which exclusion of data inputs is necessary. One possibility would be to require the exclusion of variables that do not logically relate to default—an approach that relies on intuitive decisions because we do not know what causes default. Importantly, it is hard to reconcile these intuitive decisions with the data-driven approach of machine learning, in which variables will be selected for carrying predictive power.

Furthermore, the effectiveness of these types of restrictions is called into question because even excluding other variables that are correlated with race has limited effect in big data. In the third graph, we depict the predicted default rates using a decision rule that was created by excluding race and the ten variables that most correlate with race. Despite significantly reducing the number of variables that correlate strongly with race, the disparity still persists for the three racial groups even though it is now smaller. The purpose of these three graphs is to demonstrate the impact that correlated data has on the decision rule even when we exclude the forbidden characteristics or variables that may be deemed closer to the forbidden characteristics.[32] In big data, even excluding those variables that individually relate most to the "forbidden input" does not necessarily significantly affect how much

[32]For a demonstration of the limited effect of excluding race from default prediction using data on real mortgage performance, see Fuster, et al, *Predictably Unequal* at *26–30.

pricing outputs vary with, say, race.[33]

If disparate impact is a proxy for disparate treatment or a means of enforcing disparate treatment law,[34] we may find it sufficient that we can guarantee that there is no direct effect of race on the decision. Although it has long been recognized that a disparate impact claim does not require a showing of intention to discriminate, which has traditionally been understood as the domain of disparate treatment, it is disputed whether the purpose of disparate impact is to deal with cases in which intention is hard to prove or whether the very foundation of the disparate impact doctrine is to deal with cases in which there is no intention to discriminate. There are several aspects of how disparate impact has been interpreted and applied that support the notion that it is a tool for enforcing disparate treatment law rather than a theory of discrimination that is philosophically distinct.[35] According to the Supreme Court in *Texas Department of Housing & Community Affairs v Inclusive Communities Project*,[36] "Recognition of disparate-impact liability under the FHA plays an important role in uncovering discriminatory intent: it permits plaintiffs to counteract unconscious prejudices and disguised animus that escape easy classification as disparate treatment."[37]

On the other hand, most formal articulations are clear that disparate impact can apply even when there is no discriminatory intent, not only when discriminatory intent is not

[33] An alternative approach taken in the algorithmic fairness literature is to transform variables that correlate with the forbidden characteristic as a way of "cleaning" the training data. See, for example, Feldman, et al, *Certifying and Removing Disparate Impact*, at *260 (2014). For a general discussion of these approaches, see James E. Johndrow and Kristian Lum, *An Algorithm for Removing Sensitive Information: Application to Race-Independent Recidivism Prediction* *3 (Annals of Applied Statistics, forthcoming 2018), archived at http://perma.cc/CR67-48PN.

[34] For a discussion of this view, see, for example, Richard Arneson, *Discrimination, Disparate Impact, and Theories of Justice*, in Deborah Hellman and Sophia Moreau, eds, *Philosophical Foundations of Discrimination Law* 87, 105 (Oxford 2013). See also generally Primus, 117 Harv L Rev at 493.

[35] A disparate impact claim can be sustained only if the plaintiff has not demonstrated a "business necessity" for the conduct. *Texas Department of Housing & Community Affairs v Inclusive Communities Project, Inc*, 135 S Ct 2507, 2517 (2015). Conduct that lacks a business justification and led to a discriminatory outcome raises the suspicion that it is ill-intended. Moreover, many cases that deal with human decision-making seem to imply that there may have been intent to discriminate. For example, in *Watson v Fort Worth Bank & Trust*, 487 US 977 (1988), the Court emphasized that, while the delegation of promotion decisions to supervisors may not be with discriminatory intent, it is still possible that the particular supervisors had discriminatory intent. Id at 990.

[36] 135 S Ct 2507 (2015).

[37] Id at 2511–12.

established.[38] This understanding of the disparate impact doctrine also seems more in line with perceptions of regulators and agencies that enforce antidiscrimination law in the context of credit.[39] To the extent that disparate impact plays a social role beyond acting as a proxy for disparate treatment,[40] we may not find it sufficient to formally exclude race from the data considered.

3.3.2 Exclusion May Be Undesirable

Another criterion for the exclusion of inputs beyond the forbidden characteristics themselves are variables that may be biased. Variables could be biased because of some measurement error or because the variables reflect some historical bias. For example, income may correlate with race and gender as a result of labor market discrimination, and lending histories may be a result of prior discrimination in credit markets.[41] The various ways variables can be biased has been discussed elsewhere.[42]

When data includes biased variables, it may not be desirable to exclude a protected characteristic because the inclusion of protected characteristics may allow the algorithm to correct for the biased variable.[43] For example, over the years there has been mounting criticism of credit scores because they consider measures of creditworthiness that are

---

[38]See 78 Fed Reg at 11461 ("HUD . . . has long interpreted the Act to prohibit practices with an unjustified discriminatory effect, regardless of whether there was an intent to discriminate.").

[39]See id. See also 12 CFR § 1002.6(a) ("The legislative history of the [ECOA] indicates that the Congress intended an 'effects test' concept . . . to be applicable to a creditor's determination of creditworthiness.").

[40]This approach to disparate impact has been labeled as an "affirmative action" approach to disparate impact. See Arneson, *Discrimination, Disparate Impact, and Theories of Justice* at 105–08. For further discussion of the different theories of disparate impact and their application to antidiscrimination policy in the algorithmic context see: Tal Z. Zarsky, An Analytical Challenge: Discrimination Theory in the Age of Preditive Analytics, 14 I/S: J.L. & Pol'y for Info. Soc'y 11(2018).

[41]See, for example, Hurley and Adebayo, 18 Yale J L & Tech at 156 (discussing how past exclusion from the credit may affect future exclusion through credit scores).

[42]See, for example, Barocas and Selbst, 104 Cal L Rev at 677.

[43]See generally Kleinberg, et al, 108 AEA Papers and Proceedings. See also Zachary C. Lipton, Alexandra Chouldechova, and Julian McAuley, *Does Mitigating ML's Impact Disparity Require Treatment Disparity?* *9 (ArXiv, 2018), archived at http://perma.cc/KH2Q-Z64F (discussing the shortcomings of algorithms that train on data with group membership but are group blind when used to make predictions and arguing that a transparent use of group membership can better achieve "impact parity").

more predictive for certain groups while overlooking indications of creditworthiness more prevalent for minority groups.[44]

One way this might happen is through credit rating agencies focusing on credit that comes from mainstream lenders like depository banking institutions. However, if minority borrowers are more likely to turn to finance companies that are not mainstream lenders, and this credit is treated less favorably by credit rating agencies,[45] the credit score may reflect the particular measurement method of the agency rather than underlying creditworthiness in a way that is biased against minorities. If credit scores should receive less weight for minority borrowers, a machine-learning lender that uses a credit score as one of its data inputs would want to be able to use race as another data input in order to distinguish the use of credit scores for different groups.[46]

Achieving less discriminatory outcomes by including forbidden characteristics in the prediction algorithm presents a tension between the input-focused "disparate treatment" and the outcome-focused "disparate impact" doctrines. This tension created by the requirements to ignore forbidden characteristics and yet assure that policies do not create disparate impact, thereby requiring a consideration of people's forbidden characteristics, has been debated in the past.[47] In the context of machine-learning credit pricing, including forbidden characteristics could potentially allow for the mitigation of harm from variables that suffer from biased measurement error.[48]

---

[44]See, for example, Lisa Rice and Deidre Swesnik, *Discriminatory Effects of Credit Scoring on Communities of Color*, 46 Suffolk U L Rev 935, 937 (2013) ("Credit-scoring systems in use today continue to rely upon the dual credit market that discriminates against people of color. For example, these systems penalize borrowers for using the type of credit disproportionately used by borrowers of color.").

[45]See Rice and Swesnik, 46 Suffolk U L Rev at 949.

[46]Prior economic literature on affirmative action has argued that group-blind policies may be second best in increasing opportunities for disadvantaged groups relative to group-aware policies. See, for example, Roland G. Fryer Jr and Glenn C. Loury, *Valuing Diversity*, 121 J Pol Econ 747, 773 (2013).

[47]See, for example, the discussion of Primus, 117 Harv L Rev 494, in Justice Scalia's concurrence in *Ricci v DeStefano*, 557 US 557, 594 (2009) (Scalia concurring).

[48]See generally, for example, Kleinberg, et al, 108 AEA Papers and Proceedings (applying this logic to a hypothetical algorithm to be used in college admissions; arguing that facially neutral variables like SAT score can be correlated with race for a variety of reasons, such as the ability to take a prep course; and generating a theorem showing that excluding race from consideration while leaving in variables correlated with race leads to

To summarize this Section, despite the significant opportunity for increased transparency afforded by automated pricing, legal rules that focus on input regulation will have limited effect.[49] On the one hand, unlike the human decision-making context, we can guarantee that input has been excluded. However, if we care about outcome, we should move away from focusing on input restrictions as the emphasis of antidiscrimination law. This is because input exclusion cannot eliminate and may even exacerbate pricing disparity.

3.4 Algorithmic Construction and Process-Focused Discrimination

In the context of human credit-pricing, most of the decision-making process is opaque, leading to a limited ability to examine this process. Consider the mortgage lending cases we describe in the Introduction, in which mortgage brokers determined the markup above the "par rate" set by the mortgage originator. In those cases, the broker decisions led to racial price disparity. However, it is unclear exactly why the broker decisions led to these differences. The brokers could have considered customers' race directly and charged minorities higher prices or perhaps the brokers put disproportionate weight on variables that are correlated with race, such as borrower neighborhood. Although the exact nature of these decisions could lead to different conclusions as to the discrimination norm that was violated, these questions of the exact nature of the broker decisions remain speculative given that we have no record of the decision-making process.

To overcome the inherent difficulty in recovering the exact nature of the particular decision that may have been discriminatory, cases often abstract away by focusing on the facilitation of discriminatory decisions. The limited ability to scrutinize the decisions

less equitable outcomes).

[49]In our simulation, we focus on excluding group identities when fitting the prediction function. In the context of linear regression, using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models. See also Indrė Žliobaitė and Bart Custers, *Using Sensitive Personal Data May Be Necessary for Avoiding Discrimination in Data-Driven Decision Models*, 24 Artificial Intelligence & L 183 (2016), (proposing a procedure that uses the sensitive attribute in the training data but then producing a decision rule without it); Devin G. Pope and Justin R. Sydnor, *Implementing Anti-discrimination Policies in Statistical Profiling Models*, 3 Am Econ J 206 (2011) (making a similar proposal).

themselves leads courts and regulators to identify the discretion provided to brokers when setting the mortgage terms as the conduct that caused disparity.

Algorithmic decision-making presents an opportunity for transparency. Unlike the human decision-making context in which many aspects of the decision remain highly opaque—sometimes even to the decisionmakers themselves—in the context of algorithmic decision-making, we can observe many aspects of the decision and therefore scrutinize these decisions to a greater extent. The decision process that led to a certain outcome can theoretically be recovered in the context of algorithmic decision-making, providing for potential transparency that is not possible with human decision-making.[50]

However, this transparency is constrained by the limits on interpretability of decision rules. Prior legal writing on algorithmic fairness often characterized algorithms as opaque and uninterpretable.[51] However, whether an algorithm is interpretable depends on the question being asked. Despite the opaqueness of the mortgage broker decisions, these decisions are not referred to as uninterpretable. Instead, analytical and legal tools have been developed to consider the questions that can be answered in that context. Similarly, in the context of machine learning, we need to understand what types of questions can be answered and analyzed and then develop the legal framework to evaluate these questions. There are many ways in which algorithms can be interpreted thanks to the increased replicability of their judgements. Indeed, we highlight in Section 3.5 a crucial way in which algorithms can be interpreted for the purposes of ex ante regulation.

One potential way to interpret algorithmic decisions is to consider which variables are used by the algorithm, equivalent to interpreting coefficients in regression analysis. Typically, in social science research, the purpose of regression analysis is to interpret the coefficients of the independent variables, which often reflect a causal effect of the independent variable

---

[50] This may not be true for aspects of the process that involve human discretion, such as the label and feature selection.

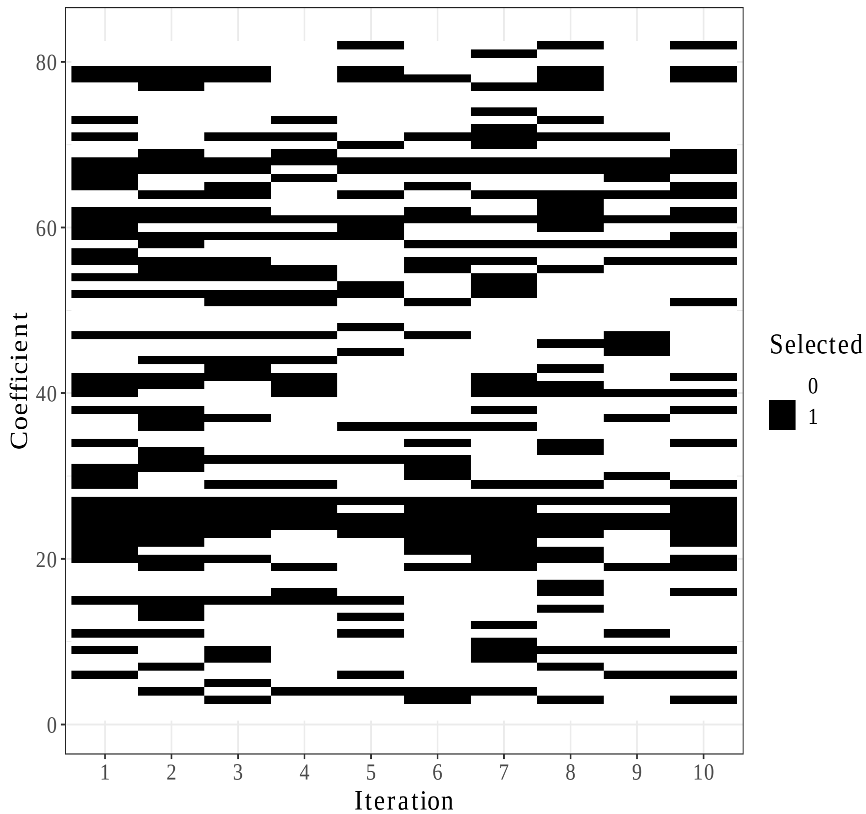
[51] See for example, Matthew Adam Bruckner, *The Promise and Perils of Algorithmic Lenders' Use of Big Data*, 93 Chi Kent L Rev 3, 44 (2018) (discussing how consumers may find it difficult to protect themselves because "many learning algorithms are thought to be quite opaque").

on the dependent variable. Analogously, in the case of machine learning, the decision rule is constructed by an algorithm, providing two related opportunities: First, the algorithm provides a decision rule (prediction function) that can be inspected and from which we can presumably determine which variables matter for the prediction. Second, we can inspect the construction of the decision rule itself and attempt to measure which variables were instrumental in forming the final rule. In the case of a prediction rule that creates differing predictions for different groups, we may want to look to the variables used to make a prediction to understand what is driving the disparate predictions.

However, in the context of machine-learning prediction algorithms, the contribution of individual variables is often hard to assess. We demonstrate the limited expressiveness of the variables an algorithm uses by running the prediction exercise in our simulation example repeatedly. Across ten draws of data from that same population, we fit a logistic lasso regression: in every draw we let the data choose which of the many characteristics to include in the model, expecting that each run should produce qualitatively similar prediction functions. Although these samples are not identical, because of the random sampling, they are drawn from the same overall population, and we therefore expect that the algorithmic decisions should produce similar outputs.

The outcome of our simulation exercise documents the problems with assessing an algorithm by the variables it uses. The specific representation of the prediction functions and which variables are used in the final decision rule vary considerably in our example. A graphic representation of this instability can be found in Figure 3.2. This Figure records which characteristics were included in the logistic lasso regressions we ran on ten draws from the population. Each column represents a draw, while the vertical axis enumerates the over eighty dummy-encoded variables in our data set. The black lines in each column reflect the particular variables that were included in the logistic lasso regression for that sample draw. While some characteristics (rows) are consistently included in the model, there are few discernible patterns, and an analysis of these prediction functions based on which variables were included would yield different conclusions from draw to draw, despite originating from similar data.

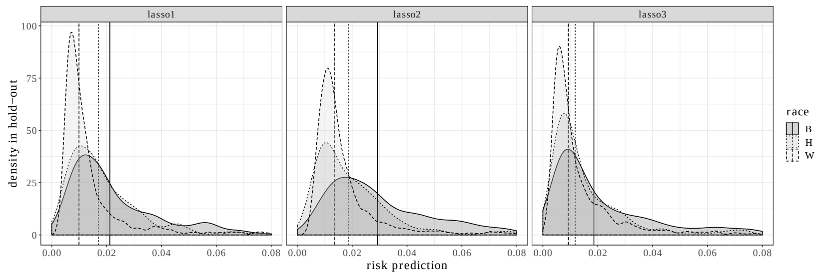
Figure 3.2: *Included predictors in a lasso regression across ten samples from the same population*



Importantly, despite these rules looking vastly different, their overall predictions indeed appear qualitatively similar. Figure 3.3 shows the distribution of default predictions by group for the first three draws of our ten random draws, documenting that they are qualitatively similar with respect to their pricing properties across groups. So while the prediction functions look very different, the underlying data, the way in which they were constructed, and the resulting price distributions are all similar.

The instability of the variables chosen for the prediction suggests that we should be skeptical about looking at the inclusion of certain variables to evaluate the process by which the decision rule was constructed and to determine the relationship between those variables and the ultimate decision. The primary object of a machine-learning algorithm is the accuracy of the prediction and not a determination of the effect of specific variables in

Figure 3.3: *Distribution of default predictions for the first three lasso predictors*



determining the outcome. When there are many possible characteristics that predictions can depend on and algorithms choose from a large, expressive class of potential prediction functions, many rules that look very different have qualitatively similar prediction properties. Which of these rules is chosen in a given draw of the data may come down to a flip of a coin. While these rules still differ in their predictions for some individuals, to the degree that we care only about the rules' overall prediction performance or overall pricing distributions, the specific representation of prediction functions may therefore not generally be a good description of relevant properties of the decisions.[52] In general, when data is high dimensional and complex machine-learning algorithms are used, a determination of conduct based on variable-importance measures is limited absent a specific notion of "importance" or interpretability that encapsulates the considerations relevant for the discrimination context.[53]

The problem with interpretability illustrated by this instability is important for how law approaches the evaluation of algorithms. We demonstrate that the deconstruction of the prediction in the hope of recovering the causes of disparity and maybe even consideration of which variables should be omitted from the algorithm to reduce disparity is limited. Even without this issue of instability of the prediction rule, it may be hard to intelligently describe

[52]See Sendhil Mullainathan and Jann Spiess, *Machine Learning: An Applied Econometric Approach*, 31 J Econ Perspectives 87, 97 (2017) ("Similar predictions can be produced using very different variables. Which variables are actually chosen depends on the specific finite sample. . . . This problem is ubiquitous in machine learning.").

[53]See Zachary C. Lipton, *The Mythos of Model Interpretability* *2 (ArXiv, 2016), archived at http://perma.cc/3JQD-6CJ4 (highlighting that there is no common notion of "interpretability" for machine-learning models because the goals of interpretation differ); Leilani H. Gilpin, et al, *Explaining Explanations: An Approach to Evaluating Interpretability of Machine Learning* (ArXiv, 2018), archived at http://perma.cc/B789-LZBL, (critically reviewing attempts to explain machine-learning models and noting the diversity of their goals).

the rule when it is constructed from many variables, all of which receive only marginal weight. Therefore, legal rules that seek to identify the cause or root of disparate decisions cannot be easily applied.

Legal doctrines that put weight on identifying a particular conduct that caused disparity will not be able to rely on variable inclusion or general-purpose importance analysis alone. Courts have interpreted the FHA, the ECOA, and their implementing regulations as requiring that the plaintiff demonstrate a causal connection between a discriminatory outcome and a specific practice in order to establish a prima facie case of discrimination.[54] The Supreme Court recently affirmed the requirement for the identification of a particular policy that caused the disparity in *Inclusive Communities*, in which it said: "[A] disparate-impact claim that relies on a statistical disparity must fail if the plaintiff cannot point to a defendant's policy or policies causing that disparity."[55] In the mortgage lending cases we discuss in the Introduction, the conduct was the discretion given to mortgage lender employees and brokers. In other housing contexts, policies that have been found to cause a discriminatory effect include landlord residency preferences that favor people with local ties over outsiders and land-use restrictions that prevent housing proposals that are of particular value to minorities.[56] At first blush, it may seem appropriate to ask which of the variables that are included in the decision rule are those that led to the pricing differential, in accordance with the conduct identification requirement of the discrimination doctrine.[57]

---

[54]See, for example, *Wards Cove Packing Co v Atonio*, 490 US 642, 657–58 (1989) (establishing a "specific causation requirement" for disparate impact claims under Title VII of the Civil Rights Act of 1964). See also 12 CFR Part 100.

[55]*Inclusive Communities*, 135 S Ct at 2523.

[56]For additional examples of challenged policies, see Robert G. Schwemm and Calvin Bradford, *Proving Disparate Impact in Fair Housing Cases after* Inclusive Communities, 19 NYU J Legis & Pub Pol 685, 718–60 (2016).

[57]Although automated credit systems have been challenged in court, court decisions rarely provide guidance on this question. For example, in *Beaulialice v Federal Home Loan Mortgage Corp*, 2007 WL 744646, *4 (MD Fla), the plaintiff challenged the automated system used to determine her eligibility for a mortgage. Although the defendant's motion for summary judgement was granted, the basis for the decision was not that the plaintiff had not demonstrated conduct for a plausible claim of disparate impact. Alternatively, if the mere decision to use an algorithm is the "conduct" that caused discrimination, the requirement will be devoid of any meaningful content, strengthening the conclusion that the identification of a policy should be replaced with a greater emphasis on other elements of the analysis. This analysis is the outcome analysis discussed in the next Section.

However, our example above demonstrates that such an analysis is questionable and unlikely to be appropriate as the algorithmic equivalent of identifying conduct absent context-specific importance and transparency measures that apply to the legal context.[58] Antidiscrimination doctrine should therefore move away from this type of abstract decision rule analysis as a central component of antidiscrimination law.

3.5 Implied Prices and Outcome-Focused Discrimination

In Section 3.4 we argue that, despite its purported transparency, the analysis of machine-learning pricing is constrained by limits to the interpretability of abstract pricing rules. In this Section, we argue that the replicability that comes with automation still has meaningful benefits for the analysis of discrimination when the pricing rule is applied to a particular population. The resulting price menu is an object that can be studied and analyzed and, therefore, should play a more central role in discrimination analysis. We consider an ex ante form of regulation that we call "discrimination stress testing," which exploits the opportunity that automated decision rules can be evaluated before they are applied to actual consumers. Our focus is on how to evaluate whether a pricing rule is fair, not on how to construct a fair pricing rule.

The final stage of a lending decision is the pricing "outcome," meaning the prices paid by consumers. In a world in which credit pricing involves mortgage brokers setting the final lending terms, pricing outcomes are not known until the actual prices have materialized for actual consumers. When pricing is automated, however, we also have information about pricing, even before customers receive loans, from inspecting the pricing rule. Furthermore, the pricing rule can be applied to any population, real or theoretical, to understand the pricing distribution that the pricing rule creates. Therefore, the set of potential outcomes based on algorithmic decision-making that a legal regime can analyze is broader than the set

[58] There may be situations in which a particular aspect of the construction of the algorithm can be identified as leading to discrimination. As discussed in prior literature, biased outcomes may be a result of human decisions regarding the use and construction of the data. See, for example, Barocas and Selbst, 104 Cal L Rev at 677–93; Hurley and Adebayo, 18 Yale J L & Tech at 173.

of outcomes that can be analyzed in the case of human decision-making, and the richness of information that is available at an earlier point in time means that the practices of the lender can be examined before waiting a period of time to observe actual prices.

Although pricing-outcome analysis plays an important conceptual role in discrimination law, it is debatable how to practically conduct this analysis. Formally, outcome analysis that shows that prices provided to different groups diverge is part of the prima facie case of disparate impact. However, despite the centrality of outcome analysis, there is surprisingly little guidance on how exactly to conduct outcome analysis for the purposes of a finding of discrimination.[59] For example, we know little about the criteria to use when comparing two consumers to determine whether they were treated differently or, in the language of the legal requirement, whether two "similarly situated" people obtained different prices.

In addition, there is little guidance on the relevant statistical test to use. As a result, output analysis in discrimination cases often focuses on simple comparisons and regression specifications[60] and then moves quite swiftly to other elements of the case that are afforded a more prominent role, such as the discussion of the particular conduct or policy that led to a disparate outcome.

In the case of machine learning, we argue that outcome analysis becomes central to the application of antidiscrimination law. As Section 3.3 and Section 3.4 discuss, both input regulation and decision process scrutiny are limited in the context of machine-learning pricing. Crucial aspects of current antidiscrimination law that focus on the procedure of creating the eventual prices are limited by the difficulties of interpreting the decision rule that leads to the disparity and the challenges of closely regulating data inputs. Therefore, antidiscrimination law will need to increase its focus on outcome analysis in the context of

---

[59]See Schwemm and Bradford, 19 NYU J Legis & Pub Pol at 690–92 (arguing that neither HUD Regulation 12 CFR Part 11 nor *Inclusive Communities*, both of which endorse discriminatory effects claims under the FHA, provide any guidance on how to establish differential pricing for a prima facie case of discrimination and showing that lower courts rarely followed the methodology established under Title VII).

[60]See Ayres, Klein, and West, *The Rise and (Potential) Fall of Disparate Impact Lending Litigation* at 236 (analyzing *In re Wells Fargo Mortgage Lending Discrimination Litigation*, 2011 WL 8960474 (ND Cal), in which plaintiffs used regression analysis to prove unjustified disparate impacts, as an example of how plaintiffs typically proceed).

machine-learning credit pricing.[61]

The type of ex ante analysis, which we call "discrimination stress testing," is most similar to bank stress testing, which also evaluates an outcome using hypothetical parameters. Introduced in February 2009 as part of the Obama Administration's Financial Stability Plan and later formalized in Dodd-Frank,[62] stress tests require certain banks to report their stability under hypothetical financial scenarios.[63] These scenarios are determined by the Federal Reserve and specify the macroeconomic variables, such as the GDP growth and housing prices, that the bank needs to assume in its predicted portfolio risk and revenue. The results of these tests help to determine whether the bank should increase its capital and provide a general assessment of the bank's resilience. This allows for a form of regulation that is forward-looking and provides a consistent estimate across banks.[64] In a discrimination stress test, the regulator would apply the pricing rule of the lender to some hypothetical population before the lender implements the rule to evaluate whether the pricing meets some criteria of disparity.[65]

Developing the precise discrimination stress test requires articulating how the test will be implemented and the criteria used to judge pricing outcomes. A full analysis of these issues is beyond the scope of this Essay. Instead, we highlight two main concerns in the development the discrimination stress test. First, we demonstrate the significance of

---

[61]Pauline Kim argues that, in the algorithmic context, employers should be allowed to rely on the "bottom-line defense," thereby recognizing an increased role for outcome-based analysis in this context. See Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 Wm & Mary L Rev 857, 923 (2016).

[62]Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub L No 111-203, 124 Stat 1376 (2010), codified at 12 USC § 1503 et seq.

[63]12 USC § 5365.

[64]See Michael S. Barr, Howell E. Jackson, and Margaret E. Tahyar, *Financial Regulation: Law and Policy* 313 (Foundation 2016).

[65]The power to announce future regulatory intent already exists within the CFPB's regulatory toolkit in the form of a No-Action Letter, through which it declares that it does not intend to recommend the initiation of action against a regulated entity for a certain period of time. For example, in September 2017, the CFPB issued a No-Action Letter to Upstart, a lender that uses nontraditional variables to predict creditworthiness, in which it announced that it had no intention to initiate enforcement or supervisory action against Upstart on the basis of ECOA. See Consumer Financial Protection Bureau, *No-Action Letter Issued to Upstart Network* (Sept 14, 2017), archived at http://perma.cc/N4SU-2PRS.

selecting a particular population to which the pricing rule is applied. Second, we discuss the importance of the particular statistical test used to evaluate pricing disparity.

3.5.1 Population Selection

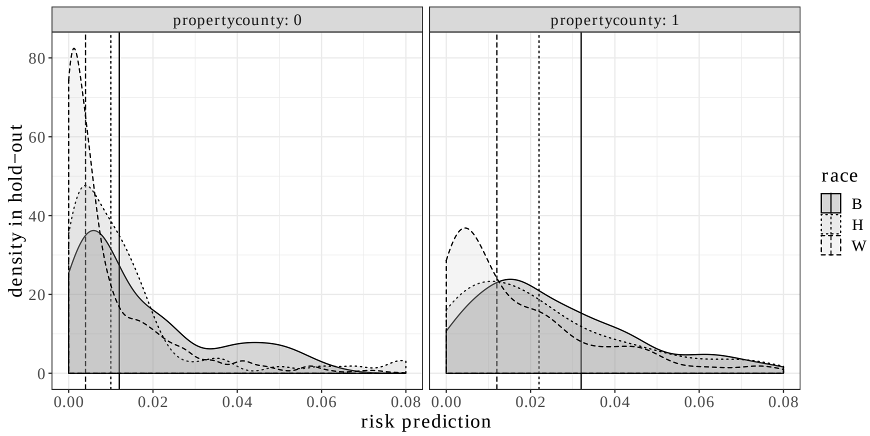
The first aspect of the discrimination stress test that we highlight is that disparity highly depends on the particular population to which the pricing rule is applied. The opportunity in the context of machine-learning pricing is that prices can be analyzed ex ante. However, this analysis can be conducted only when applying the rule to a particular population. Therefore, regulators and policy makers need to determine what population to use when applying a forward-looking test.

The decision of which population to use for testing is important because the disparity created by a pricing rule is highly sensitive to the particular population. If price disparity is created by groups having different characteristics beyond the protected characteristic, such as race, the correlations of characteristics with race will determine price disparity.

We demonstrate the sensitivity of disparate outcomes to the particular borrower population by applying the same price rule to two different populations. We split our simulated sample into two geographical groups. One group covers lenders from Suffolk County, which covers some of the more urban areas of the Boston metropolitan area, and the other group covers more rural areas. Using the same prediction rule of default, we plot the distribution by race. While the rule—in this case a prediction based on a random forest—is exactly the same, the distribution of default predictions is qualitatively different between applicants in Suffolk County (right panel of Figure 3.4) and those in more rural areas of the Boston metropolitan area (left panel). Specifically, the same rule may induce either a very similar (left) or quite different (right) distribution of predictions by group.

The sensitivity of outcomes to the selected population highlights two important considerations for policy makers. First, it suggests that regulators should be deliberate in their selection of the population to use when testing. For example, they may want to select a population in which characteristics are highly correlated with race or one that represents

Figure 3.4: *Risk predictions from the same prediction function across different neighborhoods*



more vulnerable lenders.[66] Furthermore, regulators should select the sample population based on specific regulatory goals. If, for example, regulators seek to understand the impact of a pricing rule on the specific communities in which a lender operates, regulators may select a sample population of those communities rather than a nationally representative sample. Second, if regulators wish to compare lender pricing rules, they should keep the population constant across lenders. This would provide for a comparable measure of disparity between lenders. Such meaningful comparisons are not possible with human decision-making when there is no pricing rule and only materialized prices. If regulators evaluate lending practices using ex post prices, differences between lenders may be driven by differences in decision rules or the composition of the particular population that received the loan.

The sensitivity of price disparities to the population also suggests that regulators should not disclose the exact sample they use to test discrimination. In this respect, the design of the discrimination stress test could be informed by financial institution stress testing. While the general terms of the supervisory model are made public, many of the details used to protect revenue and losses are kept confidential by regulators and are changed periodically,

[66] Regulators are often interested in who the lender actually serviced, which may reveal whether the lender was engaging in redlining or reverse redlining. Clearly, this hypothetical is inappropriate for that analysis. For further discussion of redlining and reverse redlining, see Gano, 88 U Colo L Rev at 1124–28.

limiting financial institutions' ability to game the specifics of the stress test.[67] Similarly, for discrimination stress testing, the exact data set used could be kept confidential so that lenders are not able to create decision rules that minimize disparity for the specific data set alone.

Another benefit of population selection is that it allows for price disparity testing even when lenders do not collect data on race. Although mortgage lenders are required to collect and report race data under the HMDA, other forms of lending do not have an equivalent requirement. This creates significant challenges for private and public enforcement of the ECOA, for example. Discrimination stress testing offers a solution to the problem of missing data on race. With discrimination stress testing, the lender itself would not have to collect race data for an evaluation of whether the pricing rule causes disparity as long as the population the regulator uses for the test includes protected characteristics. Because the regulator evaluates the pricing rule based on the prices provided to the hypothetical population, it can evaluate the effect on protected groups regardless of whether this data is collected by the lender.

3.5.2 Test for Disparity

Once the pricing rule is applied to a target population, the price distribution needs be evaluated. We focus on two aspects of this test: namely, the criterion by which two groups are compared and the statistical test used to conduct the comparison. A large literature originating in computer science discusses when algorithms should be considered fair.[68]

The first aspect of a disparity test is the criterion used to compare groups. For example, we could ignore any characteristics that vary between individuals and simply consider

---

[67]See Barr, Jackson, and Tahyar, *Financial Regulation: Law and Policy* at 313 ("In essence, the Federal Reserve Board, by changing the assumptions and keeping its models cloaked, is determined that its stress tests cannot be gamed by the financial sector.").

[68]See, for example, Corbett-Davies, et al, *Algorithmic Decision Making* at *2. See also generally Feldman, et al, *Certifying and Removing Disparate Impact* at *2–3. For a recent overview of the different notions of fairness, see Mark MacCarthy, *Standards of Fairness for Disparate Impact Assessment of Big Data Algorithms*, 48 Cumb L Rev 67, 89–102 (2017).

whether the price distribution is different by group. Alternatively, the criterion for comparison could deem that certain characteristics that may correlate with group membership should be controlled for when comparing between groups. When controlling for these characteristics in disparity testing, only individuals that share these characteristics are compared.[69] Courts consider this issue by asking whether "similarly situated" people from the protected and nonprotected group were treated differently.[70] Suppose that individuals with the same income, credit score, and job tenure are considered "similarly situated." The distribution of prices is then allowed to vary across groups provided that this variation represents only variation with respect to those characteristics that define "similarly situated" individuals.[71] For example, prices may still differ between Hispanic and white applicants to the degree that those differences represent differences in income, credit score, and job tenure.[72]

The longer the list of the characteristics that make people "similarly situated," the less likely it is that there will be a finding of disparity.[73] Despite the importance of this question, there is little guidance in cases and regulatory documents on which characteristics make people similarly situated.[74] An approach that considers what is predictive of being similarly

---

[69] A formalization of this idea appears in Ya'acov Ritov, Yuekai Sun, and Ruofei Zhao, *On Conditional Parity as a Notion of Non-discrimination in Machine Learning*, (ArXiv, 2017), archived at http://perma.cc/A92T-RGZW. They argue that the main notions of nondiscrimination are a form of conditional parity.

[70] See *BP Energy Co*, 828 F3d at 967. This requirement has also been referred to as the requirement that demonstration of disparate impact focus on "appropriate comparison groups." See Schwemm and Bradford, 19 NYU J Legis & Pub Pol at 698. See also Jennifer L. Peresie, *Toward a Coherent Test for Disparate Impact Discrimination*, 84 Ind L J 773, 776–79 (2009).

[71] See Cynthia Dwork, et al, *Fairness through Awareness*, Proceedings of the 3rd Innovations in Theoretical Computer Science Conference 214, 215 (2012) (providing a concept that can be seen as an implementation of "similarly situated" people being treated the same through connecting a metric of distance between people to how different their outcomes can be).

[72] See generally Robert Bartlett, et al, *Consumer-Lending Discrimination in the Era of FinTech*, *1 (UC Berkeley Public Law Research Paper, Oct 2018), archived at http://perma.cc/6BM8-DHVR. In that paper, varying credit prices and rejection rates for ethnic groups are decomposed into effects driven by "life-cycle variables" and ethnic disparities that are not driven by these variables and therefore, according to the authors, discriminatory.

[73] Ayres characterizes the problem as a determination of what variables to include as controls when regressing for the purpose of disparate impact. See Ian Ayres, *Three Tests for Measuring Unjustified Disparate Impacts in Organ Transplantation: The Problem of "Included Variable" Bias*, 48 Perspectives in Biology & Med S68, S69–70 (2005).

[74] One exception is the 1994 Policy Statement on Discrimination in Lending by HUD, the Department of

situated would mean that, by definition, the algorithm is not treating similarly situated people differently. Especially in a big data world with a large number of correlated variables, a test of statistical parity thus requires a clear implementation of similar situated to have any bite. Therefore, a determination of what makes people "similarly situated" is primarily a normative question that lawmakers and regulators should address.

The determination of who is "similarly situated" is distinct from an approach of input restriction. Restricting inputs to "similarly situated" characteristics would guarantee that there is no disparity; however, this is not necessary. Although a complete discussion of the conditions under which input variables that do not constitute "similarly situated" characteristics do not give rise to a claim of disparity is beyond the scope of this Essay, we highlight two considerations. First, as we argue throughout the Essay, the particular correlations of the training set and holdout set will affect pricing disparity, and so little can be determined from the outset. If, for example, a characteristic does not correlate with race, its inclusion in the algorithm may not lead to disparity. Second, the statistical test should include a degree of tolerance set by the regulator. When this tolerance is broader, it is more likely that characteristics included in the algorithm may not give rise to a claim of disparity, even when they are not "similarly situated" characteristics.

In addition to the criterion used to compare groups, the regulator requires a test in order to determine whether there is indeed disparity.[75] Typically, for such a test, the regulator needs to fix a tolerance level that expresses how much the distribution of risk predictions

---

Justice, and other agencies, which suggested that the characteristics listed in the HMDA do not constitute an exhaustive list of the variables that make people similarly situated. Department of Housing and Urban Development, Interagency Policy Statement on Discrimination in Lending, 59 Fed Reg 18267 (1994).

[75]The algorithmic fairness literature includes many different tests, some of which are summarized by MacCarthy, 48 Cumb L Rev at 86–89. One of the only examples of an articulated statistical test is the "four-fifths rule" adopted by the Equal Employment Opportunity Commission in 1979. 29 CFR § 1607.4(D) ("A selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded by Federal enforcement agencies as evidence of adverse impact."). We do not discuss the rule because its formulation does not seem natural in a context like credit pricing, in which there is not a single criterion with pass rates. In addition, the extent to which this test is binding is not clear given the tendency of courts to overlook it. For further discussion, see Schwemm and Bradford, 19 NYU J Legis & Pub Pol at 706–07. For an application of this test in the algorithmic fairness literature, see generally Feldman, et al, *Certifying and Removing Disparate Impact*.

may deviate across groups between similarly situated individuals.

3.6 Conclusion

In this Essay, we present a framework that connects the steps in the genesis of an algorithmic pricing decision to legal requirements developed to protect against discrimination. We argue that there is a gap between old law and new methods that can be bridged only by resolving normative legal questions. These questions have thus far received little attention because they were of less practical importance in a world in which antidiscrimination law focused on opaque human decision-making.

While algorithmic decision-making allows for pricing to become traceable, the complexity and opacity of modern machine-learning algorithms limit the applicability of existing legal antidiscrimination doctrine. Simply restricting an algorithm from using specific information, for example, would at best satisfy a narrow reading of existing legal requirements and would typically have limited bite in a world of big data. On the other hand, scrutiny of the decision process is not always feasible in the algorithmic decision-making context, suggesting a greater role for outcome analysis.

Prices set by machines also bring opportunities for effective regulation, provided that open normative questions are resolved. Our analysis highlights an important role for the statistical analysis of pricing outcomes. Because prices are set by fixed rules, discrimination stress tests are opportunities to check pricing outcomes in a controlled environment. Such tests can draw on criteria from the growing literature on algorithmic fairness, which can also illuminate the inherent tradeoffs between different notions of discrimination and fairness.

This is a watershed moment for antidiscrimination doctrine, not only because the new reality requires an adaptation of an anachronistic set of rules but because philosophical disagreements over the scope of antidiscrimination law now have practical and pressing relevance.

References

Acquisti, A., Taylor, C. and Wagman, L. (2016). The economics of privacy. *Journal of economic Literature*, **54** (2), 442–92.

Agarwal, S., Liu, C. and Souleles, N. S. (2007). The reaction of consumer spending and debt to tax rebates—evidence from consumer credit data. *Journal of political Economy*, **115** (6), 986–1019.

Aleo, M. and Svirsky, P. (2008). Foreclosure fallout: The banking industry's attack on disparate impact race discrimination claims under the fair housing act and the equal credit opportunity act. *BU Pub. Int. LJ*, **18**, 1.

Angeletos, G.-M., Laibson, D., Repetto, A., Tobacman, J. and Weinberg, S. (2001). The hyperbolic consumption model: Calibration, simulation, and empirical evaluation. *Journal of Economic perspectives*, **15** (3), 47–68.

Aridor, G., Che, Y.-K. and Salz, T. (2020). The economic consequences of data privacy regulation: Empirical evidence from gdpr. *NBER working paper*, (w26900).

Arkes, H. R., Joyner, C. A., Pezzo, M. V., Nash, J. G., Siegel-Jacobs, K. and Stone, E. (1994). The psychology of windfall gains. *Organizational behavior and human decision processes*, **59** (3), 331–347.

Ayres, I., Klein, G. and West, J. (2017). The rise and (potential) fall of disparate impact lending litigation. *evidence and innovation in housing law and policy*, **231**, 240–46.

Bakos, Y., Marotta-Wurgler, F. and Trossen, D. R. (2014). Does anyone read the fine print? consumer attention to standard-form contracts. *The Journal of Legal Studies*, **43** (1), 1–35.

Bar-Gill, O., Ben-Shahar, O. and Marotta-Wurgler, F. (2017). Searching for the common law: The quantitative approach of the restatement of consumer contracts. *U. Chi. L. Rev.*, pp. 7–35.

— and Davis, K. (2010). Empty promises. *S. Cal. L. Rev.*, **84**, 1.

Bartlett, R., Morse, A., Stanton, R. and Wallace, N. (2022). Consumer-lending discrimination in the fintech era. *Journal of Financial Economics*, **143** (1), 30–56.

Baugh, B., Ben-David, I., Park, H. and Parker, J. A. (2021). Asymmetric consumption smoothing. *American Economic Review*, **111** (1), 192–230.

Ben-Shahar, O. and Strahilevitz, L. J. (2016). Contracting over privacy: Introduction. *The Journal of Legal Studies*, **45** (S2), S1–S11.

Bergelson, V. (2003). It's personal but is it mine-toward property rights in personal information. *UC Davis L. Rev.*, **37**, 379.

Bodkin, R. (1959). Windfall income and consumption. *The American Economic Review*, **49** (4), 602–614.

Bradford, A. (2020). *The Brussels effect: How the European Union rules the world*. Oxford University Press, USA.

Calzolari, G. and Pavan, A. (2006). On the optimality of privacy in sequential contracting. *Journal of Economic theory*, **130** (1), 168–204.

Campbell, J. Y. and Mankiw, N. G. (1989). Consumption, income, and interest rates: Reinterpreting the time series evidence. *NBER macroeconomics annual*, **4**, 185–216.

Carroll, C. D. (1997). Buffer-stock saving and the life cycle/permanent income hypothesis. *The Quarterly journal of economics*, **112** (1), 1–55.

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S. and Huq, A. (2017). Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining*, pp. 797–806.

Davis, K. and Marotta-Wurgler, F. (2019). Contracting for personal data. *NYU L. Rev.*, **94**, 662–705.

Dinev, T. and Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information systems research*, **17** (1), 61–80.

Donohue, J. J. (2007). Antidiscrimination law. *Handbook of law and economics*, **2**, 1387–1472.

Dosis, A. and Sand-Zantman, W. (2019). The ownership of data. *Available at SSRN 3420680*.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O. and Zemel, R. (2012). Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*, pp. 214–226.

Feri, F., Giannetti, C. and Jentzsch, N. (2016). Disclosure of personal information under risk of privacy shocks. *Journal of Economic Behavior & Organization*, **123**, 138–148.

Frankenreiter, J. (2021). The missing 'california effect' in data privacy law. *Yale J. on Reg., forthcoming*.

— and Hermstruewer, Y. (2021). Privacy's great shock.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T. and Walther, A. (2022). Predictably unequal? the effects of machine learning on credit markets. *The Journal of Finance*, **77** (1), 5–47.

GANONG, P. and NOEL, P. (2019). Consumer spending during unemployment: Positive and normative implications. *American economic review*, **109** (7), 2383–2424.

GELMAN, M., KARIV, S., SHAPIRO, M. D., SILVERMAN, D. and TADELIS, S. (2014). Harnessing naturally occurring data to measure the response of spending to income. *Science*, **345** (6193), 212–215.

—, —, —, — and — (2020). How individuals respond to a liquidity shock: Evidence from the 2013 government shutdown. *Journal of Public Economics*, **189**, 103917.

GOMES, F., HALIASSOS, M. and RAMADORAI, T. (2021). Household finance. *Journal of Economic Literature*, **59** (3), 919–1000.

GROSSMAN, S. J. and HART, O. D. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of political economy*, **94** (4), 691–719.

HALL, R. E. (1978). Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence. *Journal of political economy*, **86** (6), 971–987.

HANN, I.-H., HUI, K.-L., LEE, S.-Y. T. and PNG, I. P. (2008). Consumer privacy and marketing avoidance: A static model. *Management science*, **54** (6), 1094–1103.

HARRIS, C. and LAIBSON, D. (2001). Dynamic choices of hyperbolic consumers. *Econometrica*, **69** (4), 935–957.

HART, O., SHLEIFER, A. and VISHNY, R. W. (1997). The proper scope of government: theory and an application to prisons. *The Quarterly Journal of Economics*, **112** (4), 1127–1161.

HASTINGS, J. and SHAPIRO, J. M. (2018). How are snap benefits spent? evidence from a retail panel. *American Economic Review*, **108** (12), 3493–3540.

HOOFNAGLE, C. J., SOLTANI, A., GOOD, N. and WAMBACH, D. J. (2012). Behavioral advertising: The offer you can't refuse. *Harv. L. & Pol'y Rev.*, **6**, 273.

HSIEH, C.-T. (2003). Do consumers react to anticipated income changes? evidence from the alaska permanent fund. *American Economic Review*, **93** (1), 397–405.

HUQ, A. Z. (2017). What is discriminatory intent. *Cornell L. Rev.*, **103**, 1211.

HURLEY, M. and ADEBAYO, J. (2016). Credit scoring in the era of big data. *Yale JL & Tech.*, **18**, 148.

JOHNSON, D. S., PARKER, J. A. and SOULELES, N. S. (2006). Household expenditure and the income tax rebates of 2001. *American Economic Review*, **96** (5), 1589–1610.

JORRING, A. (2018). The costs of financial mistakes: Evidence from us consumers. *University of Chicago Manuscript*.

KAPLAN, G. and VIOLANTE, G. L. (2014). A model of the consumption response to fiscal stimulus payments. *Econometrica*, **82** (4), 1199–1239.

—, — and Weidner, J. (2014). *The wealthy hand-to-mouth*. Tech. rep., National Bureau of Economic Research.

Karnow, C. E. (2021). The internet and contract formation. *Berkeley Bus. LJ*, **18**, 135.

Klass, G. (2019). Empiricism and privacy policies in the restatement of consumer contract law. *Yale J. on Reg.*, **36**, 45.

Kleinberg, J., Ludwig, J., Mullainathan, S. and Rambachan, A. (2018). Algorithmic fairness. In *Aea papers and proceedings*, vol. 108, pp. 22–27.

Kretschmer, M., Pennekamp, J. and Wehrle, K. (2021). Cookie banners and privacy policies: Measuring the impact of the gdpr on the web. *ACM Transactions on the Web (TWEB)*, **15** (4), 1–42.

Kueng, L. (2018). Excess sensitivity of high-income consumers. *The Quarterly Journal of Economics*, **133** (4), 1693–1751.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, **112** (2), 443–478.

Lancieri, F. (2022). Narrowing data protection's enforcement gap. *Maine L. Rev.*, **74**, forthcoming.

Lenard, T. M. and Rubin, P. H. (2010). In defense of data: Information and the costs of privacy. *Policy & Internet*, **2** (1), 149–183.

Levitin, A. J., Kim, N. S., Kunz, C. L., Linzer, P., McCoy, P. A., Moringiello, J. M., Renuart, E. A. and Willis, L. E. (2019). The faulty foundation of the draft restatement of consumer contracts. *Yale J. on Reg.*, **36**, 447.

Linden, T., Khandelwal, R., Harkous, H. and Fawaz, K. (2019). The privacy policy landscape after the gdpr.

Lipton, Z., McAuley, J. and Chouldechova, A. (2018). Does mitigating ml's impact disparity require treatment disparity? *Advances in neural information processing systems*, **31**.

Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, **16** (3), 31–57.

Litman, J. (2000). Information privacy/information property. *Stan. L. Rev.*, pp. 1283–1313.

Mangini, V., Tal, I. and Moldovan, A.-N. (2020). An empirical study on the impact of gdpr and right to be forgotten-organisations and users perspective. In *Proceedings of the 15th International Conference on Availability, Reliability and Security*, pp. 1–9.

Marotta-Wurgler, F. and Svirky, D. (2021). Do ftc privacy enforcement actions matter? compliance before and after us-eu safe harbor agreement actions.

Mastrobuoni, G. and Weinberg, M. (2009). Heterogeneity in intra-monthly consumption patterns, self-control, and savings at retirement. *American Economic Journal: Economic Policy*, **1** (2), 163–89.

Matwyshyn, A. M. (2013). Privacy, the hacker way. *S. Cal. L. Rev.*, **87**, 1.

McDonald, A. M. and Cranor, L. F. (2008). The cost of reading privacy policies. *Isjlp*, **4**, 543.

McDowall, R. A. (2019). Consumption behavior across the distribution of liquid assets.

Milberg, S. J., Burke, S. J., Smith, H. J. and Kallman, E. A. (1995). Values, personal information privacy, and regulatory approaches. *Communications of the ACM*, **38** (12), 65–74.

Mullainathan, S. and Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, **31** (2), 87–106.

Obar, J. A. and Oeldorf-Hirsch, A. (2020). The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society*, **23** (1), 128–147.

O'Donoghue, T. and Rabin, M. (1999). Doing it now or later. *American economic review*, **89** (1), 103–124.

Olafsson, A. and Pagel, M. (2018). The liquid hand-to-mouth: Evidence from personal finance management software. *The Review of Financial Studies*, **31** (11), 4398–4446.

Parker, J. A. (1999). The reaction of household consumption to predictable changes in social security taxes. *American Economic Review*, **89** (4), 959–973.

—, Souleles, N. S., Johnson, D. S. and McClelland, R. (2013). Consumer spending and the economic stimulus payments of 2008. *American Economic Review*, **103** (6), 2530–53.

Pope, D. G. and Sydnor, J. R. (2011). Implementing anti-discrimination policies in statistical profiling models. *American Economic Journal: Economic Policy*, **3** (3), 206–31.

Posner, R. A. (1981). The economics of privacy. *The American economic review*, **71** (2), 405–409.

Primus, R. (2010). The future of disparate impact. *Mich. L. Rev.*, **108** (8), 1341–1387.

Primus, R. A. (2003). Equal protection and disparate impact: Round three. *Harv. L. Rev.*, **117**, 494.

Richards, N. (2021). *Why Privacy Matters*. Oxford University Press.

Romanosky, S. and Acquisti, A. (2009). Privacy costs and personal data protection: Economic and legal perspectives. *Berkeley Tech. LJ*, **24**, 1061.

Santos, C., Nouwens, M., Toth, M., Bielova, N. and Roca, V. (2021). Consent management platforms under the gdpr: processors and/or controllers? In *Annual Privacy Forum*, Springer, pp. 47–69.

Schmitt, J., Miller, K. M. and Skiera, B. (2021). The impact of privacy laws on online user behavior. *arXiv preprint arXiv:2101.11366*.

SCHWARTZ, P. M. and SOLOVE, D. J. (2011). The pii problem: Privacy and a new concept of personally identifiable information. *NYUL rev.*, **86**, 1814.

SCHWEMM, R. G. (2015). Fair housing litigation after inclusive communities: What's new and what's not. *Colum. L. Rev. Sidebar*, **115**, 106.

SHAHAR, O. B. and STRAHILEVITZ, L. J. (2016). Contracting over privacy: Introduction. *Journal of Legal Studies*, **45**, S7.

SHAPIRO, J. M. (2005). Is there a daily discount rate? evidence from the food stamp nutrition cycle. *Journal of public Economics*, **89** (2-3), 303–325.

SHAPIRO, M. D. and SLEMROD, J. (2003). Consumer response to tax rebates. *American Economic Review*, **93** (1), 381–396.

SHEA, J. (1995). Union contracts and the life-cycle/permanent-income hypothesis. *The American Economic Review*, pp. 186–200.

SHEFRIN, H. M. and THALER, R. H. (1988). The behavioral life-cycle hypothesis. *Economic inquiry*, **26** (4), 609–643.

SLOAN, R. H. and WARNER, R. (2014). Beyond notice and choice: Privacy, norms, and consent. *J. High Tech. L.*, **14**, 370.

SOLOVE, D. and SCHWARTZ, P. M. (2019). *Privacy Law Fundamentals 2019*. IAPP.

SOLOVE, D. J. and HARTZOG, W. (2014). The ftc and the new common law of privacy. *Colum. L. Rev.*, **114**, 583–676.

SOULELES, N. S. (1999). The response of household consumption to income tax refunds. *American Economic Review*, **89** (4), 947–958.

STEPHENS JR, M. (2001). The long-run consumption effects of earnings shocks. *Review of Economics and Statistics*, **83** (1), 28–36.

STIGLER, G. J. (1980). An introduction to privacy in economics and politics. *The Journal of Legal Studies*, **9** (4), 623–644.

STRAHILEVITZ, L. J. and KUGLER, M. B. (2016). Is privacy policy language irrelevant to consumers? *The Journal of Legal Studies*, **45** (S2), S69–S95.

THALER, R. (1985). Mental accounting and consumer choice. *Marketing science*, **4** (3), 199–214.

VARIAN, H. R. (2002). Economic aspects of personal privacy. In *Cyber Policy and Economics in an Internet Age*, Springer, pp. 127–137.

VICTOR, J. M. (2013). The eu general data protection regulation: Toward a property regime for protecting data privacy. *Yale LJ*, **123**, 513.

VILJOEN, S. (2021). A relational theory of data governance. *Yale LJ*, **131**, 573.

WESTIN, A. F. (2000). Intrusions. *Public Perspective*, **11** (6), 8–11.

Zarsky, T. Z. (2013). Transparent predictions. *U. Ill. L. Rev.*, p. 1503.

— (2017). An analytic challenge: Discrimination theory in the age of predictive analytics. *ISJLP*, **14**, 11.

Zeldes, S. P. (1989). Consumption and liquidity constraints: an empirical investigation. *Journal of political economy*, **97** (2), 305–346.

Appendix A

Appendix to Chapter 1

Table A.1: *Daily Spending on Food and Drink over the Mortgage Repayment Cycle*

Sample	Full Sample			Households that have both debit and credit cards		
Cards included in spending measure	Debit and credit	Debit cards only	Credit cards only	Debit and credit	Debit cards only	Credit cards only
Days in mortgage repayment cycle						
-15 through -8	-0.010	-0.009	-0.001	-0.043*	-0.027**	-0.014
	(0.013)	(0.006)	(0.011)	(0.019)	(0.009)	(0.015)
-7 through -1	--	--	--	--	--	--
0 through +6	-0.048***	-0.039***	-0.007	-0.062**	-0.057***	-0.003
	(0.014)	(0.006)	(0.012)	(0.020)	(0.010)	(0.016)
+7 through +15	-0.028*	-0.019**	-0.009	-0.038+	-0.036***	-0.001
	(0.014)	(0.006)	(0.011)	(0.020)	(0.010)	(0.016)
Number of households	46,832	46,832	46,832	22,905	22,905	22,905
Number of observations	23,287,476	23,287,476	23,287,476	11,687,369	11,687,369	11,687,369

+ Significant at 10% level. * Significant at 5% level. ** Significant at 1% level. *** Significant at 0.1% level.

Table A.2: *Daily Spending on Food, Drink and Entertainment Away from Home as a Function of Number of Days from Nearest Mortgage Payment, by Income Tercile*

Panel A. Lowest Tercile

Sample	Full Sample			Households that have both debit and credit cards		
Cards included in spending measure	Debit and credit	Debit cards only	Credit cards only	Debit and credit	Debit cards only	Credit cards only
Days in mortgage repayment cycle						
-15 through -8	-0.012 (0.023)	-0.007 (0.011)	-0.007 (0.019)	-0.012 (0.033)	-0.028 (0.018)	0.012 (0.026)
-7 through -1	--	--	--	--	--	--
0 through +6	-0.064** (0.024)	-0.038** (0.012)	-0.021 (0.019)	-0.076* (0.036)	-0.059** (0.021)	-0.011 (0.026)
+7 through +15	-0.060* (0.024)	-0.017 (0.012)	-0.038* (0.019)	-0.047 (0.036)	-0.034+ (0.020)	-0.006 (0.027)
Number of households	15,611	15,611	15,611	7,401	7,401	7,401
Number of observations	7,284,002	7,284,002	7,284,002	3,572,840	3,572,840	3,572,840

Panel B. Mid Tercile

Sample	Full Sample			Households that have both debit and credit cards		
Cards included in spending measure	Debit and credit	Debit cards only	Credit cards only	Debit and credit	Debit cards only	Credit cards only
Days in mortgage repayment cycle						
-15 through -8	-0.052* (0.024)	-0.011 (0.011)	-0.035+ (0.019)	-0.098** (0.033)	-0.027 (0.018)	-0.060* (0.026)
-7 through -1	--	--	--	--	--	--
0 through +6	-0.088*** (0.026)	-0.055*** (0.012)	-0.023 (0.020)	-0.103** (0.036)	-0.066*** (0.019)	-0.020 (0.028)
+7 through +15	-0.070** (0.025)	-0.027* (0.012)	-0.036+ (0.020)	-0.082* (0.036)	-0.025 (0.019)	-0.050+ (0.028)
Number of households	15,611	15,611	15,611	7,967	7,967	7,967
Number of observations	7,859,910	7,859,910	7,859,910	4,109,499	4,109,499	4,109,499

Panel C. Highest Tercile

Sample	Full Sample			Households that have both debit and credit cards		
Cards included in spending measure	Debit and credit	Debit cards only	Credit cards only	Debit and credit	Debit cards only	Credit cards only
Days in mortgage repayment cycle						
-15 through -8	0.018 (0.027)	-0.014 (0.011)	0.031 (0.023)	-0.030 (0.041)	-0.031 (0.019)	0.009 (0.033)
-7 through -1	--	--	--	--	--	--
0 through +6	-0.035 (0.028)	-0.048*** (0.011)	0.017 (0.025)	-0.069+ (0.040)	-0.084*** (0.019)	0.023 (0.034)
+7 through +15	0.004 (0.027)	-0.032** (0.011)	0.035 (0.023)	-0.016 (0.041)	-0.066*** (0.019)	0.056+ (0.032)
Number of households	15,610	15,610	15,610	7,537	7,537	7,537
Number of observations	8,143,564	8,143,564	8,143,564	4,005,030	4,005,030	4,005,030

+ Significant at 10% level. * Significant at 5% level. ** Significant at 1% level. *** Significant at 0.1% level.

Table A.3: *Daily Spending on Food, Drink and Entertainment Away from Home as a Function of Number of Days from Nearest Mortgage Payment, by Principal Repayment Amount Tercile*

Panel A. Lowest Tercile

Sample	Full Sample			Households that have both debit and credit cards		
Cards included in spending measure	Debit and credit	Debit cards only	Credit cards only	Debit and credit	Debit cards only	Credit cards only
Days in mortgage repayment cycle						
-15 through -8	-0.038	-0.007	-0.033	-0.080*	-0.027	-0.048
	(0.026)	(0.010)	(0.022)	(0.037)	(0.018)	(0.030)
-7 through -1	--	--	--	--	--	--
0 through +6	-0.072**	-0.043***	-0.021	-0.089*	-0.073***	-0.007
	(0.027)	(0.011)	(0.023)	(0.040)	(0.019)	(0.031)
+7 through +15	-0.058*	-0.031**	-0.021	-0.084*	-0.048*	-0.027
	(0.027)	(0.011)	(0.022)	(0.037)	(0.019)	(0.030)
Number of households	15,611	15,611	15,611	7,477	7,477	7,477
Number of observations	7,831,422	7,831,422	7,831,422	3,835,483	3,835,483	3,835,483

Panel B. Mid Tercile

Sample	Full Sample			Households that have both debit and credit cards		
Cards included in spending measure	Debit and credit	Debit cards only	Credit cards only	Debit and credit	Debit cards only	Credit cards only
Days in mortgage repayment cycle						
-15 through -8	0.003	-0.015	0.021	-0.004	-0.035*	0.036
	(0.023)	(0.011)	(0.018)	(0.035)	(0.018)	(0.027)
-7 through -1	--	--	--	--	--	--
0 through +6	-0.026	-0.046***	0.020	-0.045	-0.058**	0.024
	(0.024)	(0.011)	(0.019)	(0.036)	(0.018)	(0.026)
+7 through +15	-0.024	-0.030**	0.009	0.000	-0.040*	0.044
	(0.024)	(0.011)	(0.018)	(0.035)	(0.018)	(0.027)
Number of households	15,611	15,611	15,611	7,416	7,416	7,416
Number of observations	7,748,167	7,748,167	7,748,167	3,782,128	3,782,128	3,782,128

Panel C. Highest Tercile

Sample	Full Sample			Households that have both debit and credit cards		
Cards included in spending measure	Debit and credit	Debit cards only	Credit cards only	Debit and credit	Debit cards only	Credit cards only
Days in mortgage repayment cycle						
-15 through -8	-0.008	-0.010	0.003	-0.060	-0.024	-0.028
	(0.027)	(0.012)	(0.022)	(0.039)	(0.020)	(0.029)
-7 through -1	--	--	--	--	--	--
0 through +6	-0.088**	-0.052***	-0.022	-0.113**	-0.079***	-0.022
	(0.028)	(0.012)	(0.022)	(0.040)	(0.020)	(0.030)
+7 through +15	-0.040	-0.014	-0.023	-0.061	-0.037+	-0.018
	(0.028)	(0.012)	(0.022)	(0.040)	(0.020)	(0.032)
Number of households	15,610	15,610	15,610	8,012	8,012	8,012
Number of observations	7,707,887	7,707,887	7,707,887	4,069,758	4,069,758	4,069,758

+ Significant at 10% level. * Significant at 5% level. ** Significant at 1% level. *** Significant at 0.1% level.

Table A.4: *Daily Spending on Food, Drink and Entertainment Away from Home as a Function of Number of Days from Nearest Mortgage Payment, by Income/Mortgage Tercile*

Panel A. Lowest Tercile

Sample	Full Sample			Households that have both debit and credit cards		
Cards included in spending measure	Debit and credit	Debit cards only	Credit cards only	Debit and credit	Debit cards only	Credit cards only
Days in mortgage repayment cycle						
-15 through -8	-0.016 (0.024)	-0.006 (0.012)	-0.009 (0.019)	-0.049 (0.034)	-0.020 (0.018)	-0.022 (0.026)
-7 through -1	--	--	--	--	--	--
0 through +6	-0.072** (0.025)	-0.037** (0.012)	-0.025 (0.020)	-0.089* (0.037)	-0.046* (0.019)	-0.026 (0.029)
+7 through +15	-0.047+ (0.025)	-0.008 (0.012)	-0.035+ (0.019)	-0.059 (0.038)	-0.024 (0.019)	-0.026 (0.028)
Number of households	15,611	15,611	15,611	7,901	7,901	7,901
Number of observations	7,523,071	7,523,071	7,523,071	3,934,746	3,934,746	3,934,746

Panel B. Mid Tercile

Sample	Full Sample			Households that have both debit and credit cards		
Cards included in spending measure	Debit and credit	Debit cards only	Credit cards only	Debit and credit	Debit cards only	Credit cards only
Days in mortgage repayment cycle						
-15 through -8	0.014 (0.024)	-0.015 (0.011)	0.032+ (0.020)	-0.008 (0.035)	-0.031+ (0.018)	0.027 (0.027)
-7 through -1	--	--	--	--	--	--
0 through +6	-0.058* (0.025)	-0.059*** (0.011)	0.006 (0.020)	-0.069+ (0.036)	-0.088*** (0.018)	0.026 (0.028)
+7 through +15	-0.015 (0.025)	-0.031** (0.011)	0.018 (0.020)	0.012 (0.036)	-0.044* (0.018)	0.056+ (0.029)
Number of households	15,611	15,611	15,611	7,515	7,515	7,515
Number of observations	7,886,497	7,886,497	7,886,497	3,897,232	3,897,232	3,897,232

Panel C. Highest Tercile

Sample	Full Sample			Households that have both debit and credit cards		
Cards included in spending measure	Debit and credit	Debit cards only	Credit cards only	Debit and credit	Debit cards only	Credit cards only
Days in mortgage repayment cycle						
-15 through -8	-0.043 (0.026)	-0.010 (0.010)	-0.033 (0.022)	-0.090* (0.038)	-0.035+ (0.019)	-0.049 (0.030)
-7 through -1	--	--	--	--	--	--
0 through +6	-0.057* (0.028)	-0.045*** (0.011)	-0.006 (0.023)	-0.093* (0.041)	-0.077*** (0.019)	-0.008 (0.032)
+7 through +15	-0.062* (0.027)	-0.036*** (0.011)	-0.020 (0.022)	-0.101 (0.039)	-0.058** (0.019)	-0.032 (0.030)
Number of households	15,610	15,610	15,610	7,489	7,489	7,489
Number of observations	7,877,908	7,877,908	7,877,908	3,855,391	3,855,391	3,855,391

+ Significant at 10% level. * Significant at 5% level. ** Significant at 1% level. *** Significant at 0.1% level.

Table A.5: *Daily Spending on Food, Drink and Entertainment Away from Home as a Function of Number of Days from Nearest Mortgage Payment, Sample Paying Credit Card Interest*

Sample	Full Sample			Households that have both debit and credit cards		
Cards included in spending measure	Debit and credit	Debit cards only	Credit cards only	Debit and credit	Debit cards only	Credit cards only
Days in mortgage repayment cycle						
-15 through -8	-0.053*	-0.028**	-0.020	-0.079**	-0.042*	-0.026
	(0.022)	(0.009)	(0.018)	(0.028)	(0.015)	(0.020)
-7 through -1	--	--	--	--	--	--
0 through +6	-0.095***	-0.066***	-0.015	-0.132***	-0.104***	-0.007
	(0.023)	(0.010)	(0.019)	(0.029)	(0.016)	(0.021)
+7 through +15	-0.083***	-0.042***	-0.032+	-0.085**	-0.066***	-0.005
	(0.023)	(0.010)	(0.018)	(0.028)	(0.015)	(0.021)
Number of households	20,301	20,301	20,301	13,560	13,560	13,560
Number of observations	10,867,250	10,867,250	10,867,250	7,102,413	7,102,413	7,102,413

+ Significant at 10% level. * Significant at 5% level. ** Significant at 1% level. *** Significant at 0.1% level.

Table A.6: *Daily Spending on Food, Drink and Entertainment Away from Home as a Function of Number of Days from Nearest Mortgage Payment, Extensive Margin*

Sample	Full Sample			Households that have both debit and credit cards		
Cards included in spending measure	Debit and credit	Debit cards only	Credit cards only	Debit and credit	Debit cards only	Credit cards only
Days in mortgage repayment cycle						
-15 through -8	-0.0001	-0.0001	-0.0001	-0.0005	-0.0004	-0.0002
	(0.0003)	(0.0002)	(0.0002)	(0.0004)	(0.0003)	(0.0003)
-7 through -1	--	--	--	--	--	--
0 through +6	-0.0014***	-0.0015***	0.0000	-0.0018***	-0.0021***	0.0001
	(0.0003)	(0.0002)	(0.0002)	(0.0004)	(0.0003)	(0.0003)
+7 through +15	-0.0011***	-0.0008***	-0.0004+	-0.0013***	-0.0011***	-0.0003
	(0.0003)	(0.0002)	(0.0002)	(0.0004)	(0.0003)	(0.0003)
Number of households	46,832	46,832	46,832	22,905	22,905	22,905
Number of observations	23,287,476	23,287,476	23,287,476	11,687,369	11,687,369	11,687,369

+ Significant at 10% level. * Significant at 5% level. ** Significant at 1% level. *** Significant at 0.1% level.

Table A.7: *Total Spending, Bank Account Balances as a Function of Number of Days from Nearest Mortgage Payment*

Panel A. Total Spending

Sample	Full Sample			Households that have both debit and credit cards		
Cards included in spending measure	Debit and credit	Debit cards only	Credit cards only	Debit and credit	Debit cards only	Credit cards only
Days in mortgage repayment cycle						
-15 through -8	-0.098 (0.178)	0.029 (0.038)	-0.050 (0.166)	0.020 (0.232)	0.007 (0.059)	0.105 (0.213)
-7 through -1	--	--	--	--	--	--
0 through +6	0.224 (0.187)	-0.267*** (0.039)	0.492** (0.175)	0.532* (0.244)	-0.370*** (0.060)	0.887*** (0.225)
+7 through +15	0.044 (0.182)	-0.078* (0.039)	0.159 (0.169)	0.176 (0.238)	-0.123* (0.060)	0.302 (0.218)
Number of households	46,832	46,832	46,832	22,905	22,905	22,905
Number of observations	23,287,476	23,287,476	23,287,476	11,687,369	11,687,369	11,687,369

+ Significant at 10% level. * Significant at 5% level. ** Significant at 1% level. *** Significant at 0.1% level.

Panel B. Bank Account Balances

Sample	Full Sample	Households that have both debit and credit cards
Days in mortgage repayment cycle		
-15 through -8	-412.66*** (25.49)	-445.85*** (27.76)
-7 through -1	--	--
0 through +6	-683.52*** (35.69)	-695.83*** (41.23)
+7 through +15	-699.83*** (34.52)	-745.43*** (42.12)
Number of households	46,602	22,891
Number of observations	22,992,156	11,618,276

+ Significant at 10% level. * Significant at 5% level. ** Significant at 1% level. *** Significant at 0.1% level.

Figure A.1: *Distribution of Fortnightly Income Arrival*

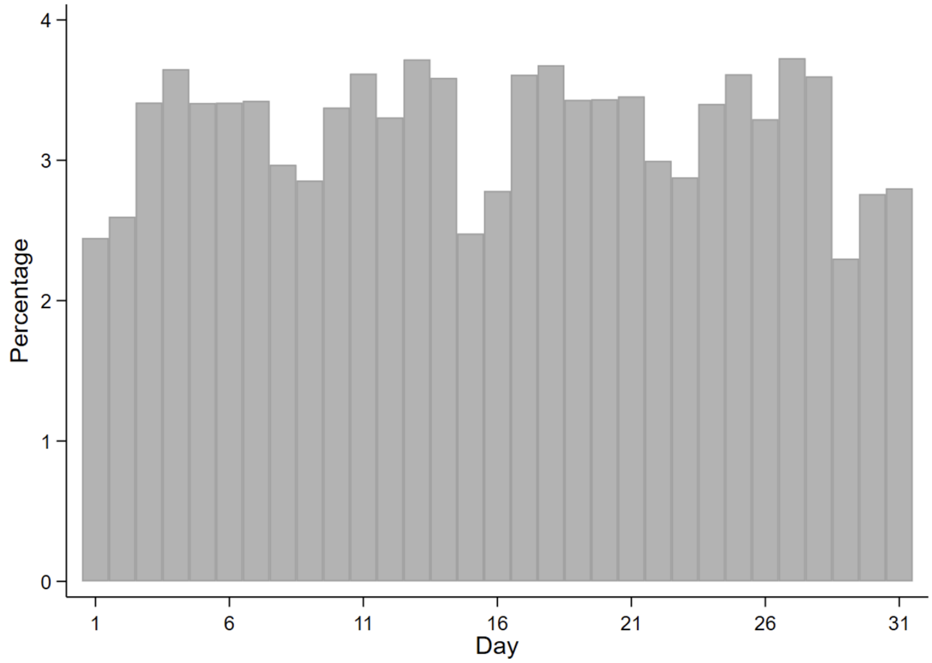
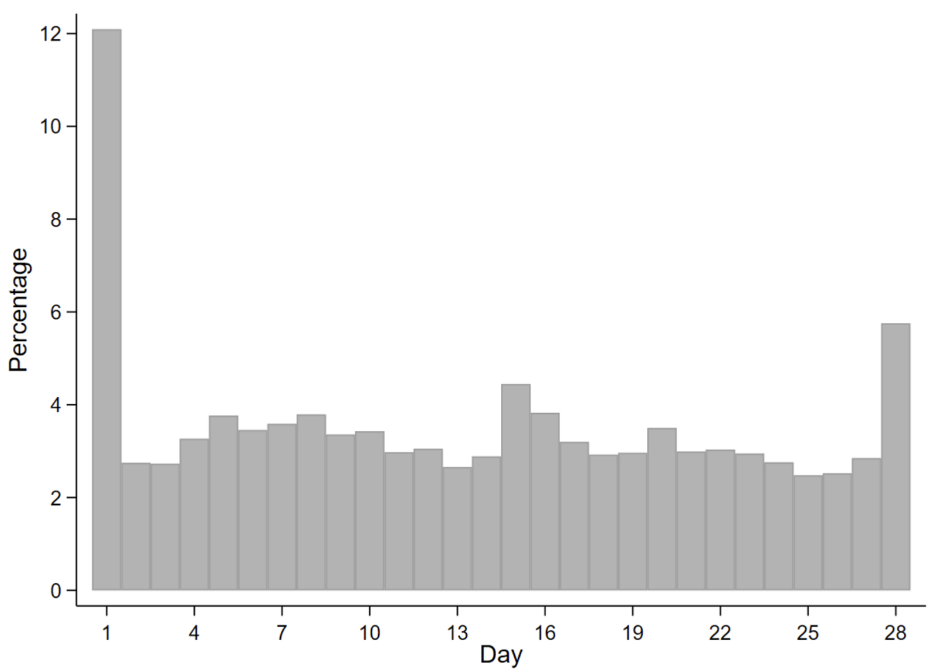


Figure A.2: *Distribution of Mortgage Repayment Dates*



Appendix B

Appendix to Chapter 2

B.1 Intensity of EU Exposure

Table B.1 shows the effect of the intensity of E.U. exposure on the likelihood of a privacy policy becoming less stringent. Unlike the main analysis of the paper, the independent variable that measures exposure to the GDPR is not a binary variable. Rather it is meant to capture the extent or intensity of E.U. exposure. As a proxy for E.U. exposure we consider the percentage of a website's traffic that comes from the E.U., as measured in October 2020.

The dependent variable in Table B.1 is a binary variable which receives a value of 0 if the policy in 2020 is unchanged from its 2018 policy with respect to the consent provision, or if the consent provision has become more stringent. The variable receives a value of 1 if the policy has a less stringent consent provision relative to the consent provision in 2018.

Table B.1 reports the estimates from a logit regression. The coefficient of the variable of percentage of EU users can be interpreted as meaning that a 1% in E.U. users leads to an increase of 5% in the likelihood of making a policy less stringent.

Table B.1: *Effect of EU Exposure Intensity*

	(1)
	consent_change_num
consent_change_num	
pct_EU_users	0.0500*
	(2.03)
_cons	-2.324***
	(-9.34)
N	238

t statistics in parentheses

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

Appendix C

Appendix to Chapter 3

C.1 Simulation Data

As discussed in the main Essay, we demonstrate our main points in a stylized simulation exercise that is calibrated to real data. Specifically, we consider a lender who prices mortgages based on an algorithmic prediction of their default risk, and we discuss the opportunities and limitations of assessing the lender's policy based on an analysis of the algorithm that yields the default predictions.

Our simulation demonstration is based on real mortgage application data from the Boston HMDA data set.[1] From this dataset we produce a simulation model that relates applicant and mortgage characteristics to the probability of default. Because mortgage defaults are not observed in this dataset, but are an essential aspect of our simulation demonstration, we impute default probabilities from loan approvals and calibrate them to overall default rates. As an important restriction to our analysis, we do not make any statements about actual defaults in this data but rather demonstrate our methodological points under this hypothesized model of default.

[1] This dataset is named after the act that required the disclosure of loan information contained in the dataset, the Home Mortgage Disclosure Act (HDMA), Pub L No 94–200, 89 Stat 1124 (1975), codified at 12 USC §§ 2801 et seq). The Boston HMDA dataset combines data from mortgage applications made in 1990 in the Boston area with a follow-up survey collected by the Federal Reserve Bank of Boston. See Alicia H. Munnell, et al, *Mortgage Lending in Boston: Interpreting HMDA Data*, 86 Am Econ Rev 25, 25–26 (1996). Our simulation is based on the full sample of 2,925 black, Hispanic, and white applicants.

Specifically, we fit a ridge-penalized logistic regression model of loan approval on approximately fifty characteristics of the loan and the borrower (including demographics, geographic information, and credit history), excluding race and ethnicity, which we then recalibrate such that the default rate among those approved for the loan matches the rate reported in *Predictably Unequal? The Effects of Machine Learning on Credit Markets*[2] from the matched HMDA-McDash dataset (Table 2). As a result, for every individual in the Boston HMDA dataset, we obtain a probability of default.

We draw samples from our simulation population as follows. First, we draw a bootstrap sample, without replacement, from the full Boston HMDA dataset. Second, for every individual in the bootstrap sample, we simulate whether that individual defaults based on the default probability implied by our calibrated simulation model. As a result, we obtain default indicators along with individual characteristics for each individual in the sample.

In our simulation demonstration, the firm constructs a prediction of default based on a training sample of two thousand consumers drawn randomly. The firm utilizes a machine-learning algorithm that uses this data to produce a prediction function that relates available consumer characteristics (potentially including race) to the predicted probability of default. We then assess the properties of a given prediction rule on a new sample of two thousand consumers.

As an example of an algorithm that produces such a prediction rule, the firm could run a simple logistic regression in their training sample that produces a prediction function of the form:

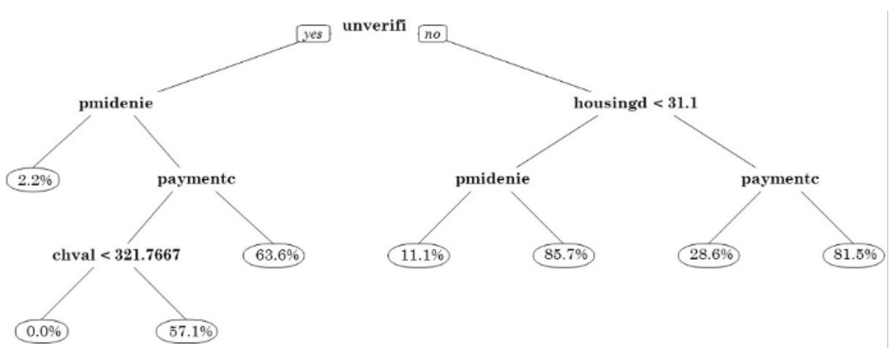
$$predicted\ probability\ of\ default\ = logistic(\alpha + \beta_1 characteristic_1 + \beta_2 characteristic_2 +\ \dots)$$

where the characteristics could be the applicant's income or credit score. While the machine-learning algorithms we consider in this Essay also produce functions that relate

[2]See Andreas Fuster, et al, *Predictably Unequal? The Effects of Machine Learning on Credit Markets* *18 (2018), archived at http://perma.cc/LYY5-SAG2.

characteristics to the probability of default, they typically take more complex forms that allow, among other things, for interactions between two or more characteristics to affect the predicted probability and are thus better suited to represent richer, possibly nonlinear relationships between characteristics and default. Some of these algorithms build on top of another simple prediction function, namely a decision (or regression) tree. The decision tree decides at every node, based on the value of one of the characteristics, whether to go left or right (for example, if income is below some threshold, go left, otherwise right), before arriving at a terminal node that returns a prediction of the probability of default of all individuals with the relevant characteristics. An example of a decision tree is given in Figure 1. Using this decision tree, the firm would predict that an individual who obtained mortgage insurance (top level, go left) but has a debt-to-income ratio of above 75 percent will have an 80 percent probability of default.

Figure C.1: *A decision tree that predicts default probability on simulated data*

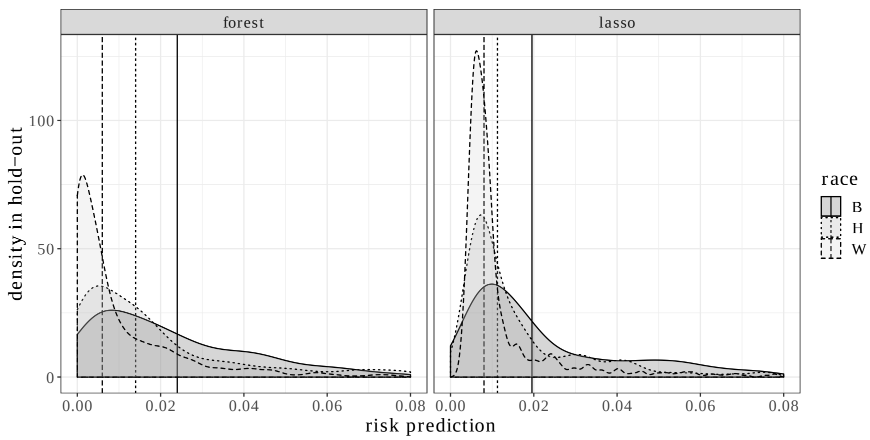


In order to analyze default predictions by group—for which we focus specifically on racial/ethnic groups in the simulation demonstration—we consider their distribution in the new ("holdout") sample of two thousand consumers drawn from the population.[3] As an example, Figure 2 plots the distribution of prices of two types of rules obtained from

[3]By analyzing the prediction rule in a new sample, we both acknowledge that the firm will use the predictions to price new clients and avoid the issue that the property of a prediction function on the data on which it was fitted can be a poor indication of its performance on new data.

machine-learning algorithms, namely a random forest[4] and a regularized logit regression.[5]

Figure C.2: *Distribution of predictions (prices) for different prediction functions*



[4]A collection of many decision trees that are averaged.

[5]A logistic regression with a lasso penalty term that improves prediction performance in high dimensions by reducing the number of characteristics entering the prediction function.