

Choice Architecture to Improve Financial Decision Making*

Jeffrey Carpenter[†] Emiliano Huet-Vaughn[‡] Peter Hans Matthews[§]

Andrea Robbett[¶] Dustin Beckett^{||} Julian Jamison^{**}

Abstract

We exploit the principles of choice architecture to evaluate interventions in the market for reloadable prepaid cards. Participants are randomized into three card menu presentation treatments - the market status quo, a regulation-inspired reform, or an enhanced reform designed to minimize attribute overload - and offered choices based on prior structural estimation of individual preferences. Consumers routinely choose incorrectly under the status quo, with tentative evidence the regulation-inspired presentation may increase best card choice, and clear evidence the enhanced reform reduces worst card choice. Welfare analysis suggests the regulation-inspired presentation offers modest gains, while the enhanced policy generates substantial benefits.

JEL Classification: C90, D04, D18, D91, G20.

Keywords: consumer finance; prepaid card; choice architecture; attribute overload; experiment; structural estimation; time preference; risk preference; convex time budget; welfare.

*We thank seminar audiences for valuable feedback at the Society for the Advancement of Behavioral Economics session at the 2017 AEA Annual Meeting, the 2017 Economic Science Association World Meeting, the 2016 New England Experimental Economics Workshop, and the Consumer Financial Protection Bureau Office of Research Seminar. We also thank our thoughtful and careful reviewers. We are grateful to the Consumer Financial Protection Bureau (CFPB) for financial support. The views expressed are those of the authors and do not necessarily reflect those of the Consumer Financial Protection Bureau.

[†]Middlebury College and IZA, jpc@middlebury.edu

[‡]Pomona College, emiliano.huet-vaughn@pomona.edu

[§]Middlebury College and Aalto University School of Business, pmatthew@middlebury.edu

[¶]Middlebury College, arobbett@middlebury.edu

^{||}Consumer Financial Protection Bureau, dustin.beckett@cfpb.gov

^{**}University of Exeter, julison@gmail.com

I Introduction

Until recently, it has been taken as given that increasing the number of choices can only improve consumer welfare. Economists are now appreciating the many instances where choice can be time-consuming, confusing, frustrating and, at times, debilitating (Scheibehenne, Greifender, and Todd, 2010; Besedes et al., 2015). Relatedly, just as a sufficiently large choice set can lead to consumer choice overload, when presented with a set of options that each vary on many different attributes, there is potential for consumers to exhibit “attribute overload,” and it is reasonable to expect they may choose incorrectly in this context too (Fasolo et al., 2007). Because the welfare consequences of consumer decision errors can be large, firms and governments have begun to consult the behavioral expertise of “choice architects” - scholars and practitioners who understand that the selection made by consumers depends to a great extent on how the choice is presented (Thaler and Sunstein, 2008).

There is by now a sizable literature on the phenomena of choice overload in economics and related disciplines, including considerable debate about its practical relevance (Scheibehenne, Greifender, and Todd, 2010). In this paper, we expand the literature in three distinct directions. First, we extend the understanding of the “paradox of choice” (Schwartz, 2004; Besedes et al, 2012) in the context of consumer goods and services. Specifically, we find that consumers experience overload, making more welfare-diminishing choices, when the number of attributes per choice is increased, even when the size and all other relevant features of the choice set are held constant. Because our experimental design customizes treatments at the individual level, using structurally estimated utility parameters derived from individuals’ pre-treatment choices, it provides a broader understanding of mistakes - that is, not just dominated choices, a first in the literature. Further, and critically, it allows us to precisely calculate the welfare loss of mistakes in ways that would otherwise not be possible, and therefore extends the litera-

ture on overload (Iyengar and Lepper, 2000; Lee and Lee, 2004), however defined. Put differently, we are able to provide a new understanding of *what* constitutes a mistake, and, *how much* that mistake costs.

Second, consistent with the increased influence of choice architects in both the private and public sectors, we embed the problem of attribute overload within the broader framework of information architecture in decision-making (Bertrand and Morse, 2011; Besedes et al., 2015) and design potential interventions accordingly. Samek et al (2016) and Peters et al (2009) have previously shown the importance of information architecture (in their case, the interaction with information using visual and other tools) in influencing decisions in multi-attribute environments. More proximately, Agnew and Szykman (2005) show that the effects of information architecture - in particular, the adoption of one of the “tabular” formats we experimentally assign - varies with financial literacy, something that we are able to control for here.

Third, we situate our research in an important but understudied consumer finance market where choice architects are set to play an increasingly influential regulatory role: the market for prepaid cards. Soll et al (2013), which evaluated whether the architecture prescribed in the CARD Act of 2009 increased consumer understanding of traditional credit cards, is in this sense a natural antecedent for our work in a related but distinct consumer financial product market.

Over the past decade, the open-loop prepaid card market in the United States has grown from \$14 billion to \$295 billion (in 2016), an annual rate of growth of more than 30%, and is expected to reach an estimated \$352.6 billion by 2020 (Rhine et al., 2007; Greene et al., 2016; Consumer Reports, 2017; Sloane, 2017).¹ A dominant segment of

¹Prepaid cards differ from debit cards “in that a debit card draws on an account owned by the cardholder” and they differ from credit cards in that “they do not draw on a line of credit established in the name of the cardholder.” Rather, they are payment devices where funds are loaded on an account associated with the payment device (not the cardholder). Open-loop cards “connect to one of the major payment network rails to complete transactions” (e.g. Visa, Mastercard, Discover, or American Express) (Sloane, 2017).

this market, accounting for a third of the current total, is the general purpose reloadable (GPR) card.² Consumers use GPRs as they would other forms of payment, to make purchases at merchants, to pay bills online, to obtain cash from ATM machines, and even to receive their pay from employers. GPRs tend to be used by the un- or under-banked, consumers who typically do not have (or cannot get) a bank account. These people are disproportionately likely to be poor, young, female, divorced or separated and living in neighborhoods with higher levels of violent crime (Hayashi and Cuddy, 2014; Greene and Shy, 2015; Cole and Greene, 2016; Ratcliffe et al., 2017). On the one hand, GPRs can be both more accessible and convenient to some consumers because one does not need to have a bank account to obtain one, but, on the other hand, as with other financial products (Choi et al., 2010), they are often associated with high fees. The fees can be numerous and their purpose - and more importantly, how to avoid them - is often unclear. The list of fees includes, but is not limited to, “periodic fees,” which are fees for holding a card for a specific time period, “per purchase fees,” “cash reload fees,” “ATM balance inquiry fees,” “customer service fees”, and even “inactivity fees.” In fact, the typical card carries between 7 and 15 fees (Pew, 2012).

Because the “choice problem” of the un-banked often does not include whether or not to use GPRs (given lack of access to traditional banking tools, GPRs are often relied on as a last resort), it is the task of regulators to ensure that the GPR market is not exploitative. While conventional wisdom now holds that consumers who overdraw their bank accounts often are likely to be better off using GPRs, the fees associated with GPRs make them very expensive ways to store and access their funds (Hayashi and Cuddy, 2015).³ Moreover, the current presentation of the fees associated with GPRs varies from provider to provider and many of the fee details are buried in the small print or online. From a choice architecture perspective, this design of the fee disclosures is poor

²The term general purpose is used because, unlike some other prepaid cards, they are not limited to specific uses, such as travel or within a specific store.

³In a survey of consumers, a Pew Foundation survey (2014) found that “avoiding overdrafts” and “avoiding spending more money than they have” were two of the top four reasons to acquire a card.

if, of course, the object is to facilitate the choices that best reflect consumer preferences (CFPB, 2014(a)).

In 2016, the Consumer Financial Protection Bureau (CFPB) finalized a rule to, among other things, create tailored provisions governing disclosures for prepaid accounts (including GPRs). Under the rule, financial institutions would be required to provide consumers with certain disclosures before acquiring the prepaid account. The “short form” disclosure presents a subset of a card’s fees, terms, and conditions, in a format that is intended to be easy for consumers to read, and to facilitate comparisons across prepaid products. The “long form” is an exhaustive list of all of a card’s fees and qualifying conditions, among other information, presented in a tabular format. In retail environments, the new rule requires financial institutions to prominently display the short form disclosure on the card’s packaging,⁴ while the long form disclosure may be provided after acquisition as long as it is made available via telephone and via the world wide web prior to acquisition.⁵

The CFPB designed the short form disclosures in a process of iterative testing and design.⁶ In doing so, the CFPB considered the lessons of the choice architecture literature (e.g., Thaler et al., 2012 or Johnson et al., 2012; Peters et al., 2009). In particular, CFPB was guided by the principles of “parsimony,” “comparability,” and “evaluability” (Johnson et al., 2012). The principle of parsimony led to the two-tiered design (short form disclosure and long form disclosure), with the short form intended to highlight a subset of the prepaid account’s most important attributes. The principle of comparability led to a rule that will standardize prepaid disclosures across products and across acquisition channels (e.g., retail and online purchases), and a simple, tabular design. The principal

⁴The short forms provided by the CFPB in its rule are technically “model forms,” and financial institutions must therefore provide a disclosure that is “substantially similar” to the CFPB’s short form disclosure.

⁵These requirements vary depending on the acquisition channel. Generally, both the short form disclosure and long form disclosure must be provided prior to account acquisition. However, there are exceptions in certain cases, such as acquisition through the retail channel.

⁶See ICF International, 2014; 2015.

of evaluability led the CFPB toward a testing plan in which form designs were iterated over five rounds of consumer testing to improve participants’ comprehension of fees and terms, abilities to compare across multiple products, and performance on hypothetical shopping tasks.⁷

The CFPB regulation of the GPR market - finally implemented in the spring of 2019 - offers an interesting real-world setting to further examine the principles of choice architecture on a large scale. This analysis forms part of our paper. Our main contribution, however, comes from the tighter identification and measurement offered by a controlled laboratory experiment we ran, in the spirit of Smith (1994), to evaluate the consequences of the policy intervention before it took place.

Our experiment took place in two parts. During the first part, we collected incentivized risk and time preferences along with demographics and data on cognitive ability and financial literacy. In the second part, which took place seven weeks later, the participants came back to the lab to do more, related, preference and individual-decision making work. For the most part, however, the second stage comprised a series of “filler tasks” that obscured its true purpose, namely, the choice of a GPR. At the beginning of this stage, participants were told that to simplify the compensation process they would be paid using GPRs, making our lab experiment one where the choice of interest, though made in a lab, has a degree of realism more common to natural field experiments (akin to the real effort lab literature, Carpenter and Huet-Vaughn, 2017).

For this second-stage payment, each participant was asked to choose from among three GPRs offered, the parameters of which were individualized for each participant

⁷The CFPB also tested a decision-aid in the form of price thermometers designed to give consumers a sense of how a product’s fees compared to other prepaid cards’ fees. Test participants found these aids confusing rather than helpful and so they were abandoned in subsequent CFPB designs. In addition, CFPB considered, but did not test, breaking the prepaid account information down into “grades” (e.g., Peters et al., 2009). However CFPB observed through focus groups that consumers used prepaid accounts in a large variety of ways and that use-patterns varied considerably across individuals. These facts led the CFPB to pursue other paths since the development of grades that could parsimoniously describe prepaid cards’ fees in a way that would benefit a sufficiently large proportion of consumers presented large challenges.

(unknown to them) based on the preferences that they revealed in the first stage of the experiment.⁸ The attributes (i.e., the fees) of one card, the “anchor card,” were assigned randomly from predetermined intervals that were representative of fees incurred using the typical card in this market. A second card, the “dominated card,” had the same overall net fee structure as the anchor card (though generated from different values of the specific fees) but offered loss-prevention insurance on worse terms.⁹ The third card was constructed, based on the individual’s preferences, such that its predicted expected utility had a lower bound for the 95% forecast interval set at the upper bound of the 95% forecast interval of the anchor card. With few exceptions, in other words, the third card should be the “best card” for the consumer, given her elicited preferences, while the “dominated card” is always the worst card.

In the presentation of the three card choice menu, we randomized subjects into one of three choice architecture treatments. The first treatment was intended to be representative of some of the products in the existing GPR industry, and, thus, a baseline, or status quo, control. Here all the fees were listed in the disclosure describing the card, but they were stated as text in the “fine print.” The second treatment was designed in the spirit of the above-mentioned CFPB short form guidelines, which aim to present information clearly and concisely, in a tabular format, and in a way that facilitates comparing terms across products. In addition, we designed an alternative, enhanced disclosure that was meant to go a step further than the CFPB short form. Here, along with the attribute information being clearly presented in tabular form to enhance the comparability of the cards offered, our disclosure did some of the required arithmetic for

⁸For a methodology in the same spirit - though in a quite different context and with a wholly different purpose to our own - see the contemporaneous work of Andreoni et al., 2016, which also uses individual preference estimates to predict an individual’s preference ordering over out-of-sample choices generated by researchers.

⁹One could view the focus on loss insurance - as opposed to deposit insurance or fraud protection - as a reflection of its salience in this particular market, consistent with Consumer Reports’ reminder to readers to activate it, and with its recommendation that CFPB mandate uniformity in coverage terms. Alternatively, and following the suggestion of one reviewer, one could instead treat it as representative of all those fees whose value to the card user is uncertain.

the consumer, with the goal of increasing the parsimony of the disclosure and reducing the attribute space to the extent possible while keeping the utility associated with each choice the same. Though not limiting for our experiment, note that our enhanced disclosure would require additional regulation, if implemented. That is, to go beyond the imposition of a sensible architecture on the disclosure information in this market, and, to take the further step of enhancing that architecture to reduce attribute overload, would be costly, as it would likely necessitate the standardization of the potential fees, themselves, in addition to the standardization of how those fees are disclosed.

As noted, a unique feature of our approach is the customization of individual GPR treatments based on the prior elicitation of preferences. In particular, we used our first stage data to generate structural estimates of the parameters of a standard utility function at the individual level, and then tailored the GPRs each consumer faced such that all participants saw personalized but (roughly) equivalent comparisons. This stands in marked contrast to standard protocols in which the choices are common to all participants and classification is impossible. Our reliance on individual-level preferences also allows us to conduct a more robust welfare analysis. That is, in addition to knowing whether consumers have made the “correct” choice, as measured by their elicited preferences, we can assess the implications of incorrect - and not just dominated - choices, and are able to ask whether the benefits of an intervention that improves consumer choices outweigh its possible costs.

Our results reveal that the majority of consumers choose the “wrong” GPR in the control treatment, as judged by their own preference elicitations of the optimal choice, including approximately 20% of our participants who chose the dominated card. With respect to the second treatment, in the spirit of the CFPB short form, there is a slight reduction (4pp) in the fraction choosing the dominated card while there is a 16 percentage point increase in the fraction choosing the best card; however, these differences, while large, are not significant at conventional levels. In the third treatment, in which the

fee information is further consolidated, we find a statistically significant reduction in the fraction of consumers who choose the dominated card compared to both the control treatment (18pp) and the regulation-inspired intervention (13pp), but there is no evidence of significant movement to the best card.

These results suggest that, given the menu of choices we provided, the cost of attribute overload is born mostly by those inclined to make a very bad choice rather than by those who fail to make the best choice in our setting. In addition to reducing the likelihood of an obvious error (i.e., picking a dominated card), the third treatment also dramatically reduces the amount of time needed to make a choice. Taking this into account along with the preferences of our participants, we find suggestive evidence that the second treatment, designed in the spirit of the CFPB short form, may increase welfare modestly, and stronger evidence that the enhanced intervention, increases welfare by about a fifth.

We proceed by describing the details of our experimental design, including our preference elicitation methods, the choice architectures implemented and the other information we gathered from our participants. Our discussion of the results begins by assessing how long it took to pick a GPR, on average, and the treatment differences in this outcome. We then discuss the likelihood of picking the dominated and best cards, in turn. The last outcome that we consider is a measure of consumer welfare enabled by our methodology. Here we assess the ability of the treatments to increase the fraction of the available dollar welfare gains achieved in the experiment. We end our discussion of the main results with a few robustness checks. We then consider the external validity of our experimental findings for choices made outside of the lab, including a study of the consequences of the CFPB policy change. In the last section of the paper we summarize our findings and discuss their implications.

II Experimental Design

The experiment was conducted in two phases, which were separated by seven weeks. Both phases took place in the lab and were programmed using the experimental software z-Tree (Fischbacher 2007). The purpose of the first phase was to elicit participants’ risk and time preferences, as well as to collect background data on demographics, cognitive ability, personality, and financial literacy and attitudes. The same participants were invited back to the lab for the second phase, during which they made the choice of primary interest: selecting one of three general purpose reloadable cards. Upon entering the second stage, each participant was randomly assigned to one of three alternative experimental treatments (unbeknownst to them), each differing in how the attributes of the GPR card were presented. Participants received the GPR card of their choice at the end of the second phase. One hundred and twenty-nine participants completed both phases of the experiment and are included in our main analysis.¹⁰

II.1 Preference Elicitation

Participants’ risk and time preferences were elicited using both the Convex Time Budget (CTB) method developed by Andreoni and Sprenger (2012) and the Double Multiple Price List (DMPL) method developed by Andersen et al. (2008). In each CTB choice, participants were asked to allocate 100 tokens across two time periods. The earlier time period was always the following day, while the later time period was k days later, where $k \in \{7, 14, 21, 28, 35\}$.¹¹ Each token allocated to the later time period always paid

¹⁰An additional 36 participants completed only the first phase and did not return for the second stage. As the experimental treatment took place upon entry to the second stage, attrition from stage 1 to stage 2 is not a concern for internal validity. An additional 11 participants completed both stages but could not be included in the analysis because their preference elicitation responses do not enable us to determine a “best” card option for them. Similarly, one individual who participated in stage 2 was not included in the analysis because we could not reliably match them to a stage 1 choices due to an error in how their ID was recorded.

¹¹We chose not to offer the possibility of immediate payment for practical design reasons, and because previous work (Augenblick et al., 2015) indicates a limited role for present bias in cases, such as ours,

the participant \$0.20 on that day, while tokens allocated to the earlier period variously paid \$0.21, \$0.20, \$0.19, \$0.18, \$0.16, \$0.14, or \$0.12. Participants thus faced five time horizons, with seven relative prices in each, for a total of 35 choices.

To allocate their tokens, participants dragged a slider between the two time periods. As they moved the sliders, their screens updated dynamically with the dollar amounts they would receive on each date, as well as color-coded bars showing the proportion of tokens allocated to both periods. Figure 1 displays the seven choices presented to participants for the three week time horizon.¹² In addition to the CTB choices, participants also made binary choices between consumption in two time periods. They did this for five distinct price lists with equivalent time horizons and relative consumption magnitudes to the CTB choices (see Appendix Figure A2 for an example of one of the time price lists). Finally, all participants completed two Holt and Laury (2002) multiple price list elicitation to gather risk attitudes (see Appendix Figure A3 for an example of the risk price list). The risk and time preferences elicited using these price lists comprise our DMPL data.

At the end of the session, one decision that the participant had made was randomly selected for payment. We follow the payment protocol developed by Andreoni and Sprenger (2012). Specifically, participants were given two envelopes, labeled “tomorrow” and “later,” which they self-addressed. They were told that they would receive the “tomorrow” envelope in their campus mailbox the following day with \$5 plus any money due to them that day (based on their randomly selected decision) and would receive the “later” envelope with \$5 plus any money due to them later, k days after that.

of CTB choices over money (and it is only for identification of a hyperbolic discounting parameter that immediate payments are needed). Regarding the practical design reasons, as emphasized by Andreoni and Sprenger (2012), it is important to ensure that the method of payment is consistent across time periods, which meant that our participants collected all of their payments from their campus mailbox. The earliest they could do so was the following day. Additionally, we wanted the time periods to align with the second phase of the experiment, when the earliest they could access funds on their GPR card was the day after the experiment.

¹²We use CTB estimates of the curvature of the utility function to infer individual risk preferences, but discuss the more direct DMPL measures in the robustness section.

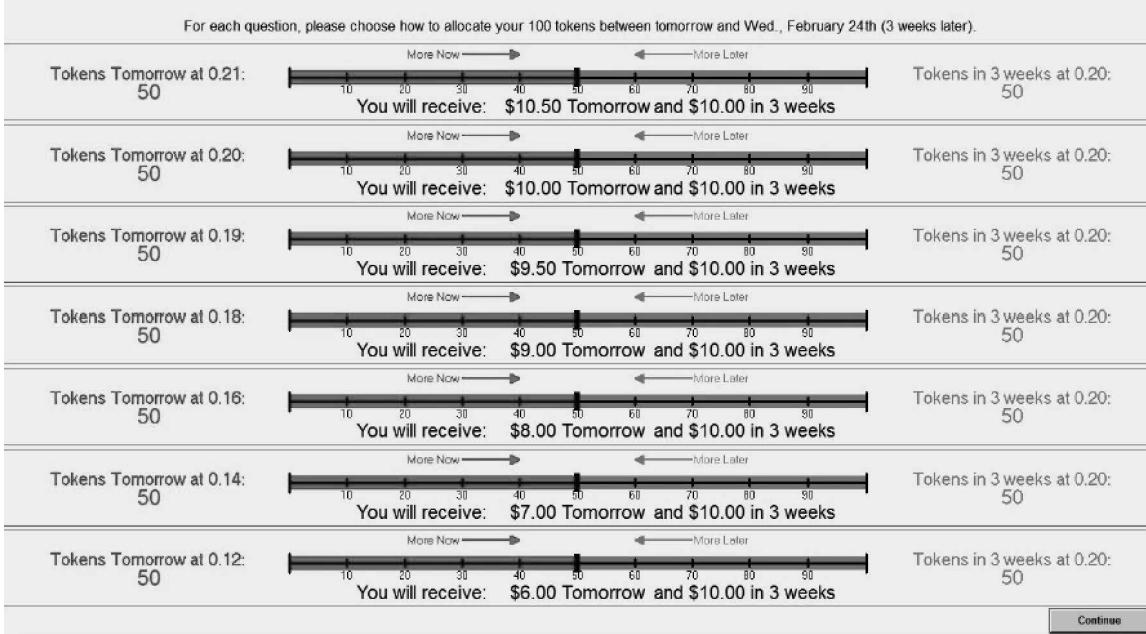


Figure 1: Example of Convex Time Budget Choices

The payment of at least \$5 in both time periods, regardless of whether the participant allocated any tokens to that time period, ensures that there are no additional transaction costs from spreading the tokens over time. The receipt of the sooner payment the following day, and in the same manner as the later payment, ensures that there is no additional risk to allocating money to the later time period.

Following the completion of the first phase, we estimated the parameters of the CRRA utility function $U(c_1, c_{1+k}) = \frac{1}{\alpha}(c_1)^\alpha + \delta^k \frac{1}{\alpha}(c_{1+k})^\alpha$ at the individual level using non-linear least squares, where c_1 is consumption the day after the experiment, c_{1+k} is consumption k days later, and α and δ are, respectively his or her curvature, or risk parameter, and discount factor. One thing to note is that because we account for the consumption delay in days, δ is the individual's *daily* discount factor implying most estimates will be between 0.99 and 1. Using this estimation method and the CTB choice data, parameter estimates of α and δ could be generated for 137 individuals (77.4% of those who completed the first phase) who regularly split their 100 tokens across both time periods. Most of the remainder exhibited behavior consistent with the maximization of

total payments over time and for these, we assume $\alpha = 1$ and $\delta = 1$. In the end, we achieved significant variation in the preferences revealed. Despite the limited possible differences in the daily discount factors and the “money maximizers,” there are 97 unique α s and 98 unique δ s and the coefficients of variation are 14% and 12%, respectively.

II.2 Card Choice

The second phase of the experiment took place seven weeks later, after all payments from the first phase had been made.¹³ At the start of the session, participants were informed that they would be paid for their participation with a GPR card (Figure A1 in the appendix). They were told that the cards enabled us to pay them over multiple time periods, while avoiding the cumbersome process of depositing envelopes in their mailboxes that they had experienced after the first phase – which was both true and provided a plausible reason for the use of the cards.¹⁴ The cards were purchased from a large gift card vendor, which offers branded debit cards to corporate clients, and were printed with our lab logo. The cards could be used to make purchases anywhere that Visa cards are accepted and were valid for six months after the experiment. We paid all of the cards’ actual fees directly to the vendor, so that they were not experienced by our participants. Instead, from the perspective of the participants, the only attributes of the cards were ones that we defined and manipulated, as described below. In addition, the product we chose did not have any fees that depended on user behavior, such as ATM withdrawals or customer service calls, which we would not have been able to control.

The cards could vary along eight attribute dimensions. Each card offered an *ini-*

¹³The 7-week window is imposed to satisfy the protocol of the CTB design while ensuring the stage 2 choices are not contaminated by stage 1 CTB payments being (differentially) received during stage 2. Since it was possible (based on chance and subject choices) for some subjects to be paid in the sixth week after stage 1 on the basis of the 35-day window choice in stage 1, we started stage 2 the week after to make sure all subjects were paid fully for stage 1 before stage 2 started.

¹⁴In addition, participants completed a filler task after making their card choice, which included similar questions to the initial survey, to reduce the likelihood of participants speculating (and possibly discussing) that the card choice was the main purpose of the second phase.

tial deposit, to be loaded onto the participant’s card the day after the experiment and accessible the same day. An additional, *reload deposit* was made four weeks later. In addition, there were five separate fees, which were modeled after actual fees typical to GPR cards, and which had the effect of reducing either the initial deposit, the reload deposit, or both. The specific fees that we added on our cards were: *activation fee* (subtracted from initial deposit), *reload fee* (subtracted from reload deposit), *monthly fee* (subtracted from both deposits), *service fee* (subtracted from both deposits), and *administration fee* (subtracted from both deposits). The final attribute concerned the riskiness of the card, which we varied using the terms of the *loss insurance*: without insurance, there was a 50% chance that \$10 would be subtracted from the reload deposit, but, with insurance, participants paid a fee to reduce the likelihood of this loss.¹⁵ The \$10 loss is in line with what many GPR cardholders are charged for a replacement card. Note that all of these attributes simply affect the magnitude and/or likelihood of the participants’ earnings in the two time periods, and thus, based on their responses from phase one of the experiment, we can estimate the expected utility that each individual participant derives from any card/bundle of attributes.

Each participant was presented a menu of three cards, which were tailored specifically for them, based on our estimates of their utility function.¹⁶ The participant was presented with three card options, which we will term Best, Middle, and Worst.¹⁷ For each participant, the attributes for the Middle card were chosen by randomly generating integer values between \$28 and \$38 for the initial and reload deposits and fee amounts between \$1 and \$5 in five cent increments. The insurance policy on the Middle card was always set such that the participant paid \$4.25 to reduce the chance of a \$10 loss to 10%

¹⁵To be clear, this is not actual FDIC insurance, but rather an experimental intervention designed to emulate it or similar vendor offered insurances.

¹⁶Of course, this assumes that the elicited preferences of our participants were stable across the intervening period, a supposition we base on the recent results of Meier and Sprenger (2015) and Schildberg-Hörisch (2018).

¹⁷In the experiment, the cards were referred to as A, B, and C and presented in a random order on the screen.

(down from 50%), a policy that would be attractive only for risk-averse participants. The attributes for the Worst card were then generated so that it was dominated by the Middle card. Specifically, the deposit in each time period, minus the associated fees, totaled the same amount as the participant’s Middle card, but the insurance policy was less favorable, only reducing the chance of loss to 15%.

Finally, the Best card attributes were generated to produce an option with higher expected utility for the participant, based on our estimates of that participant’s utility function. The Best card never came with insurance and thus was riskier than the other two cards. As a first pass, we generated attribute values for the Best card such that the upper bound of the 95% confidence interval around the expected utility of the Middle card just touched the lower bound of the confidence interval around the expected utility of the Best card. This produced Best card attribute values for 108 participants (61% of our sample) and, for the average participant, the difference in expected utility between the Best and Middle card was 4.65 (equivalent to \$4.86 on average). This method could not be used to produce attributes for participants whose preferences were consistent with money-maximization, since their parameter values were estimated with certainty, or to produce attributes within the required range for participants whose preference parameters were very imprecisely estimated. Instead, we generated Best cards for these individuals by randomly generating attributes that produced a “utility bump” of 4.65 over their Middle card.¹⁸

¹⁸Three individuals did not receive attributes using this method, due to their extremely low α values (0.499, 0.52, and 0.1). They were given the attribute values for the individual with the lowest α for whom we could generate in-range attributes (0.539). For nine remaining individuals, we were unable to generate Best card estimates using either method, because they had negative α estimates or chose all corner allocations that were not consistent with money-maximization. We gave these individuals card choices similar to those for money-maximizers, and, assigned them randomly across the three treatments. We do not consider these dozen exceptional cases in the analysis as there is little relation between their actual preference parameters and the cards they were offered (or a meaningful individualized ranking of card values).

II.3 Treatments

Three separate choice architecture treatments were conducted, which differed in how the card attributes were presented to the participants. Each participant was randomly assigned to one of the three treatments and was unaware of the other treatments. In all three treatments, the three card offers were displayed simultaneously on the screen and participants had five minutes to make a choice.¹⁹ In the Text treatment, the two deposit amounts and the insurance policy are clearly presented to participants. However, the additional fees that would cut into the participants' earnings are buried in fine print at the bottom of the card description. This treatment thus aligns with some of the least transparent products available in the GPR card market at the time of the experiment. Figure 2 presents the actual choice screen seen by a participant in the Text treatment.

The Tabular treatment is identical to the Text treatment, except that the fees are presented to the participants in a tabular format. Thus, participants in the Tabular treatment can easily compare each fee across the three cards options. Figure 3 presents a choice seen by a participant in the Tabular treatment. For each individual card, this treatment is consistent with a standardized fee disclosure template. Looking across the three cards on the screen, the tabular format resembles those used by many online “shopping comparison” tools.

The Consolidated treatment allows for an assessment of whether spreading fees across a large number of attributes impairs consumers' ability to choose the best available product. In this treatment, all of the hidden fees are summed and subtracted from the card's initial and reload amounts. The consumer therefore needs to consider only three distinct attributes in this treatment, instead of the eight attributes present in the other treatments. By eliminating attributes that simply affect the magnitude of the initial and

¹⁹After five minutes the screen started flashing in order to induce a decision, but it never actually timed out and this did not end up being a binding constraint. Subjects could use the calculators on their phones if they wished, although anecdotally this did not occur often.

Remaining Time [sec]: 295

Please click on the card you'd like to receive for your participation today. Which ever card you pick will be given to you before you leave

Card A	Card B	Card C
<p>Initial Amount \$33.00 <small>received tomorrow</small></p> <p>Reload Amount \$28.00 <small>received four weeks from tomorrow</small></p> <p>Loss Insurance Cost \$4.25 <small>There is a 50% chance that \$10 of the reload amount will be lost. This card's loss insurance reduces the chance to 10%. The insurance cost, \$4.25, will be deducted from the reload amount.</small></p> <p><small>Note: the fees listed below will be deducted from your initial amount, reload amount, or both. An activation fee of \$4.05 will be deducted from the initial amount. A monthly fee of \$3.10 will be deducted from the initial amount and the reload amount. A service fee of \$1.65 will be deducted from the initial amount and the reload amount. An administration fee of \$1.95 will be deducted from the initial amount and the reload amount. A reload fee of \$3.05 will be deducted from the reload amount.</small></p>	<p>Initial Amount \$32.00 <small>received tomorrow</small></p> <p>Reload Amount \$33.00 <small>received four weeks from tomorrow</small></p> <p>Loss Insurance Cost N/A <small>There is a 50% chance that \$10 of the reload amount will be lost. This card does not provide loss insurance.</small></p> <p><small>Note: the fees listed below will be deducted from your initial amount, reload amount, or both. An activation fee of \$1.40 will be deducted from the initial amount. A monthly fee of \$3.25 will be deducted from the initial amount and the reload amount. A service fee of \$2.75 will be deducted from the initial amount and the reload amount. An administration fee of \$2.40 will be deducted from the initial amount and the reload amount. A reload fee of \$1.90 will be deducted from the reload amount.</small></p>	<p>Initial Amount \$31.00 <small>received tomorrow</small></p> <p>Reload Amount \$26.00 <small>received four weeks from tomorrow</small></p> <p>Loss Insurance Cost \$4.25 <small>There is a 50% chance that \$10 of the reload amount will be lost. This card's loss insurance reduces the chance to 15%. The insurance cost, \$4.25, will be deducted from the reload amount.</small></p> <p><small>Note: the fees listed below will be deducted from your initial amount, reload amount, or both. An activation fee of \$3.55 will be deducted from the initial amount. A monthly fee of \$1.15 will be deducted from the initial amount and the reload amount. A service fee of \$1.45 will be deducted from the initial amount and the reload amount. An administration fee of \$2.60 will be deducted from the initial amount and the reload amount. A reload fee of \$2.55 will be deducted from the reload amount.</small></p>

Please choose your gift card by clicking on Card A, B, or C above

Figure 2: Example of Participant Card Choice in the Text Treatment

Remaining Time [sec]: 295

Please click on the card you'd like to receive for your participation today. Which ever card you pick will be given to you before you leave

Card A		Card B		Card C	
Initial Amount	\$33.00	Initial Amount	\$32.00	Initial Amount	\$31.00
<small>received tomorrow</small>		<small>received tomorrow</small>		<small>received tomorrow</small>	
Reload Amount	\$28.00	Reload Amount	\$33.00	Reload Amount	\$26.00
<small>received four weeks from tomorrow</small>		<small>received four weeks from tomorrow</small>		<small>received four weeks from tomorrow</small>	
Activation Fee	\$4.05	Activation Fee	\$1.40	Activation Fee	\$3.55
<small>deducted from initial amount</small>		<small>deducted from initial amount</small>		<small>deducted from initial amount</small>	
Monthly Fee	\$3.10	Monthly Fee	\$3.25	Monthly Fee	\$1.15
<small>deducted from initial and reload amounts</small>		<small>deducted from initial and reload amounts</small>		<small>deducted from initial and reload amounts</small>	
Loss Insurance Cost	\$4.25	Loss Insurance Cost	N/A	Loss Insurance Cost	\$4.25
<small>There is a 50% chance that \$10 of the reload amount will be lost. This card's loss insurance reduces the chance to 10%. The insurance cost, \$4.25, will be deducted from the reload amount.</small>		<small>There is a 50% chance that \$10 of the reload amount will be lost. This card does not provide loss insurance.</small>		<small>There is a 50% chance that \$10 of the reload amount will be lost. This card's loss insurance reduces the chance to 15%. The insurance cost, \$4.25, will be deducted from the reload amount.</small>	
Service Fee	\$1.65	Service Fee	\$2.75	Service Fee	\$1.45
<small>deducted from initial and reload amounts</small>		<small>deducted from initial and reload amounts</small>		<small>deducted from initial and reload amounts</small>	
Administration Fee	\$1.95	Administration Fee	\$2.40	Administration Fee	\$2.60
<small>deducted from initial and reload amounts</small>		<small>deducted from initial and reload amounts</small>		<small>deducted from initial and reload amounts</small>	
Reload Fee	\$3.05	Reload Fee	\$1.90	Reload Fee	\$2.55
<small>deducted from reload amount</small>		<small>deducted from reload amount</small>		<small>deducted from reload amount</small>	

Please choose your gift card by clicking on Card A, B, or C above

Figure 3: Example of Participant Card Choice in the Tabular Treatment

reload amounts, but don't hold any implicit value for participants, we are able to vary the number of attributes without otherwise influencing the underlying expected utility of any of the card options.²⁰ Figure 4 presents a choice seen by a participant in the Consolidated treatment.

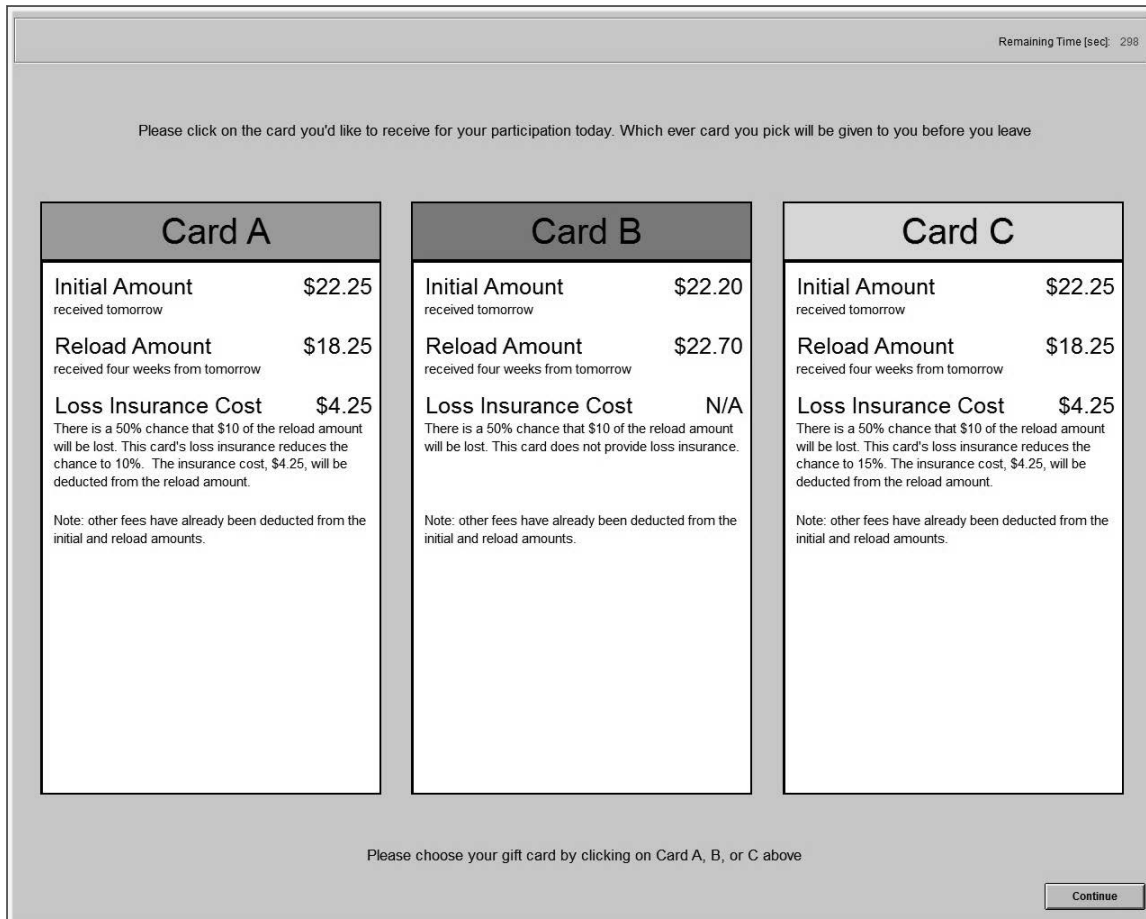


Figure 4: Example of Participant Card Choice in the Consolidated Treatment

²⁰We note that some readers may feel the application of the term “attribute” here raises philosophical or conceptual questions as to what exactly constitutes an attribute. Many of the attributes in our context can be reduced, or consolidated, to a natural common base category (e.g. period 1 dollars). In one sense, such a base category is an attribute “primitive” while the the objects that compose it are not. While this is a reasonable way to define terms, in the prepaid card market it is still the case that different fees are marketed as distinct attributes of the product (service fee; reload fee; etc.) despite their reducibility to a common base category in dollar units. Given our policy context, in the paper we think it makes sense to use attribute in this sense (and as it is used in the marketplace we study) - that is, as something that is distinguished by the seller as a unique and independent feature of the product. However, we note that attribute overload may behave differently when the cognitive limitation is forced by the difficulty of reducing “naturally” reducible attributes versus the difficulty of converting less-clearly-reducible attributes to a common scale.

One thing to consider about the Consolidated treatment is that, while reducing attributes by collapsing fees is easy to implement in the lab with the cards we have designed, in the market, collapsing fees would be more difficult without further regulation. Our MiddExLab cards incur fees at only two moments in time, when the initial amount is loaded and when the cards are reloaded. By contrast, in the market, as it is now, fees are incurred at many more distinct times so unless regulations are added to limit when fees can be charged, it would be hard to collapse attributes as comprehensively as we do. The broader point, however, is that the lab is the perfect environment for testing the effect of attribute overload, but transferring that knowledge to the marketplace would be more difficult because there cards are (currently) more complex and varied.

III Results

To begin, we present summary statistics on each of the observable variables and parameters that we hypothesize may influence card choice in the experiment, and assess whether participants are balanced across our three treatments on these measures. Summary statistics for each treatment are presented in Table 1. First, we consider family income and, cutting the sample roughly at the median, we classify participants with a family income greater than \$100 thousand as “High Income.” The next variable we consider is Cognitive Reflection, the tendency to not follow one’s “gut” and reflect more on a problem. Cognitive Reflection is measured using the standard three item test (Frederick, 2005), classifying participants as “High CRT” if they correctly answered two or three of the questions correctly. To assess financial literacy, participants answered a ten item assessment developed by Knoll and Houts (2012). Utility Curvature and Discount Factor are the participants’ preference parameters, which were estimated from the CTB data using non-linear least squares, as described in the previous section. Finally, we also consider how much is at stake for the participants when choosing among the three

Table 1: Participant Characteristics and Balance

	Text	Tabular	Consolidated
High Income ($Y_i^{Family} > \$100k$)	0.532	0.529	0.568
High CRT ($CRT_i > 1$)	0.468	0.549	0.614
High Financial Literacy ($FL_i > 7$)	0.446	0.470	0.386
Utility Curvature (α_i)	0.993	0.986	0.984
Discount Factor (δ_i)	0.999	0.999	0.998
Stakes $U^{-1}(EU_{Best} - EU_{Dominated})$	2.560	2.378	2.022

Notes: Characteristic means reported, except for preference parameters where the medians are reported.

cards. More precisely, we define Stakes to be the dollar value of the difference in expected utility between the best card and dominated card that each participant saw.²¹ On the question of balance, we find no significant pairwise differences across treatments on any of these six measures and therefore conclude that random assignment to treatment was successful.

III.1 Time Spent Deciding

We first consider whether the choice architecture affects how long the participants require to choose among the three GPR cards. The time, in seconds, that the participants spent viewing the cards before making a selection is presented in Figure 5. First, we see that participants in the Text treatment spend over two minutes, on average, evaluating the GPR card options. When the information is presented in tabular form, participants require approximately six seconds less to reach a decision, but the difference is not statistically significant ($t = 0.62$, $p = 0.535$ in a two-sided t-test). When the attributes are consolidated, however, participants require significantly less time to select a card – roughly two-thirds of the amount of time required in the Text treatment. The time spent making a choice is significantly less in the Consolidated treatment compared to the Text

²¹As estimates of utility differences will naturally depend on the first-stage estimates of the time and risk preferences that we used to formulate the card choices, to provide some degree of independence, we use the DMPL, not CTB, parameters to calculate the Stakes variable in Table 1.

($t = 3.37, p = 0.001$) and Tabular treatments ($t = 3.12, p = 0.002$). Naturally it is not always welfare-improving to spend less time making decisions, since rushing could lead to poor outcomes. In this case the Consolidated treatment appears to help along both dimensions on average, but that needn't always be the case.

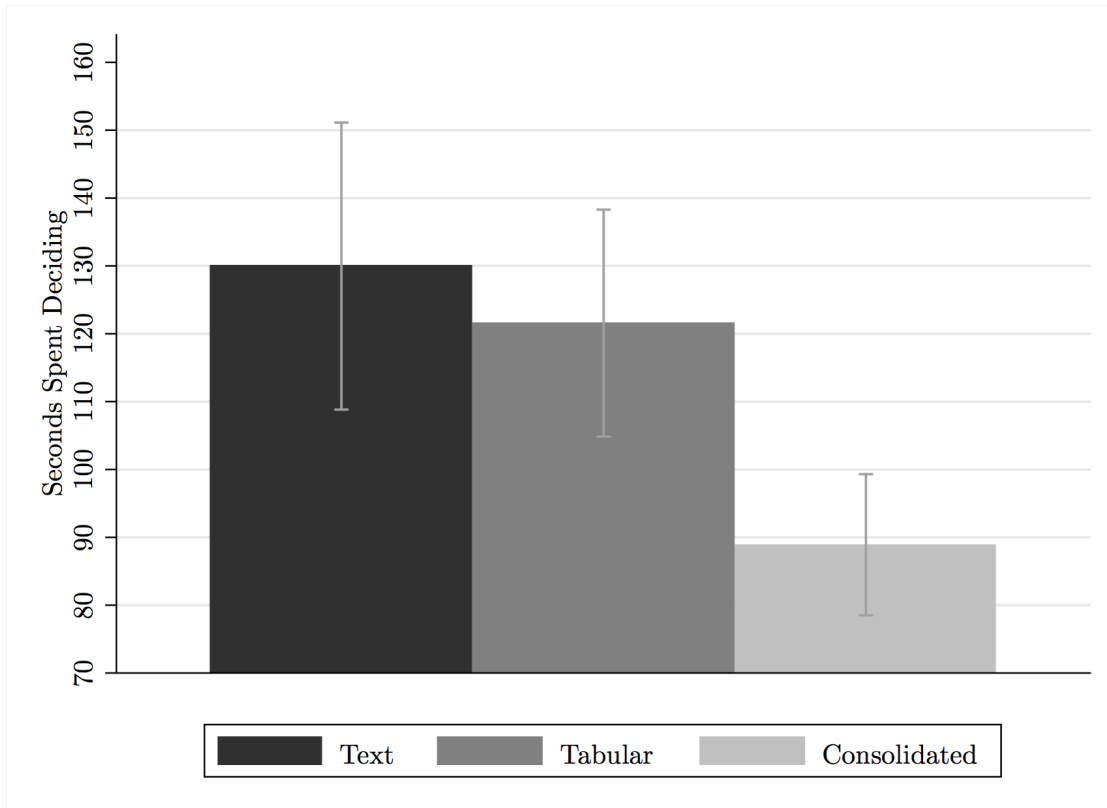


Figure 5: The time spent by participants (and the 95% C.I.s).

These findings are formalized in Table 2, which presents OLS and negative binomial regression estimates of the time spent making a decision on indicators for the Tabular and Consolidated treatments (with Text as the omitted condition). In all specifications, we find those in the Consolidated treatment make their card selection in significantly less time than those in either the Tabular or Text treatments, which themselves are not significantly different from each other, consistent with the visual evidence in Figure 5. Columns (2) and (4) include the background controls described in Table 1: High Income, High CRT, and High Financial Literacy, as well as the Stakes of the decision.

We note that the High Income participants spend significantly less time on their decision – around 28 seconds less – while those with greater financial literacy spend significantly more time. One’s score on the Cognitive Reflection Test matters in the expected direction (those exhibiting greater cognitive reflection spend more time considering their options), but the difference is not statistically significant. As can be seen across columns of Table 2, the results are nearly identical when we use the negative binomial regressor instead.

Table 2: Time Spent Deciding on GPR/prepaid Card Choice

	(1)	(2)	(3)	(4)
	OLS	OLS	NBinomial	NBinomial
Tabular	-1.775 (13.798)	-4.903 (13.235)	-1.618 (12.473)	-3.296 (11.764)
Consolidated	-38.643*** (11.704)	-37.468*** (11.631)	-41.708*** (12.390)	-37.519*** (11.994)
Stakes (in \$)		-1.750 (1.230)		-1.697 (1.350)
High Income (I)		-28.536*** (10.648)		-24.406** (9.541)
High CRT (I)		6.547 (9.945)		4.597 (9.581)
High Fin. Lit. (I)		27.996*** (10.534)		24.309** (9.772)
Constant	126.969*** (10.468)	132.325*** (13.334)		
Tabular - Consolidated	36.868*** (10.402)	31.873*** (10.718)	40.090*** (11.293)	33.707*** (11.384)
Observations	129	129	129	129
R^2 or pseudo R^2	0.084	0.183	0.010	0.019

Notes: The dependent variable is the number of seconds spent on the GPR/prepaid card choice. Robust standard errors are reported in parenthesis, and the negative binomial specifications report marginal effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

III.2 Dominance Violations

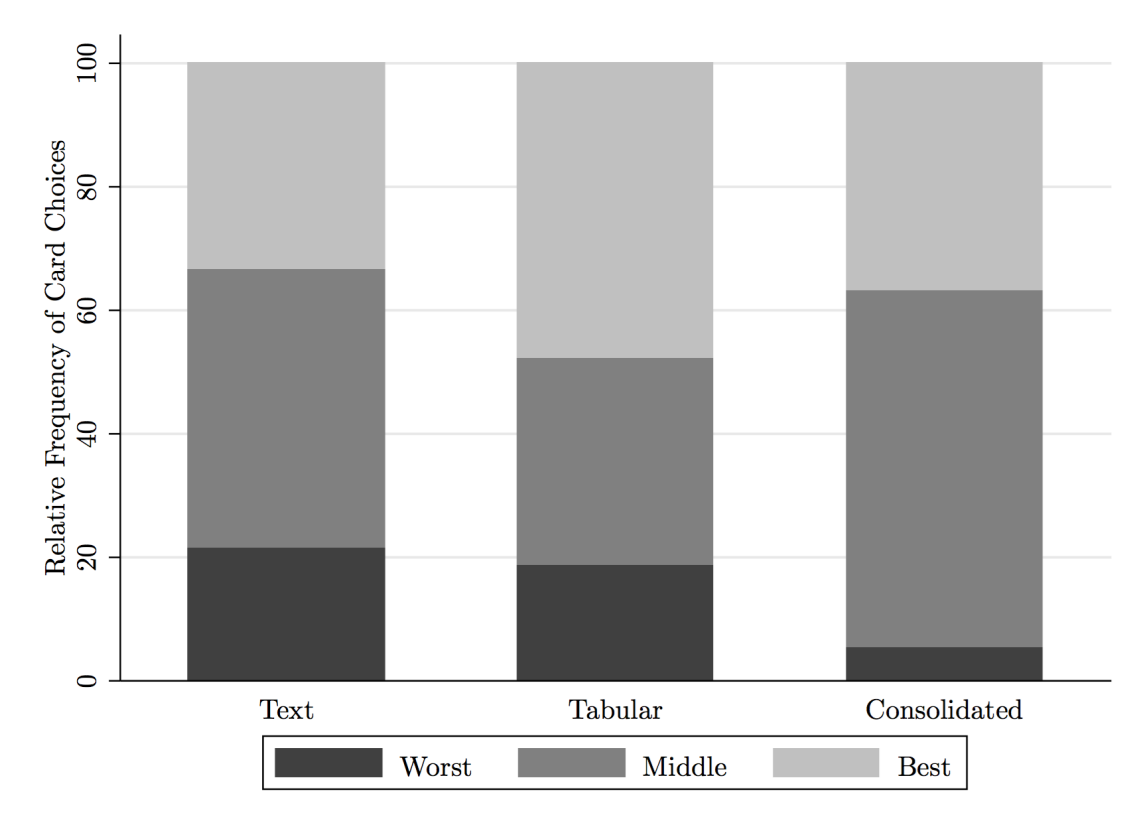


Figure 6: Frequency with which participants chose the Best, Middle, and Worst cards in each treatment.

We next consider the effect of presentation and number of attributes on the likelihood that the dominated (Worst) card is chosen. Since the Worst card is dominated by the Middle card for all participants, by design, this assessment does not depend on our preference elicitation, and serves as a non-structural method of assessing potential welfare gains from the choice architecture interventions. The relative frequency of choices across the three treatments is presented in Figure 6. In the Text treatment, where the fees are buried in the fine print (see Figure 2), the dominated Worst card is chosen more than 23% of the time. In the Tabular treatment, where participants face the same number of attributes presented with greater clarity, the likelihood of choosing the dominated card decreases only slightly, to just under 19%. In the Consolidated treatment, where

participants face far fewer attributes, the likelihood that the dominated card is chosen decreases substantially, to just 5% percent of the time.

Table 3: Choosing the Worst Card: Attribute Overload Evidence

	(1)	(2)	(3)	(4)
	OLS	OLS	Probit	Probit
Tabular	-0.045 (0.087)	-0.042 (0.090)	-0.037 (0.070)	-0.031 (0.070)
Consolidated	-0.180** (0.075)	-0.177** (0.079)	-0.208** (0.091)	-0.208** (0.089)
Stakes (in EU)		-0.010 (0.008)		-0.014 (0.011)
High Income (I)		-0.047 (0.068)		-0.041 (0.062)
High CRT (I)		-0.029 (0.070)		-0.041 (0.068)
High Fin. Lit. (I)		-0.030 (0.070)		-0.028 (0.068)
Constant	0.233*** (0.065)	0.307*** (0.091)		
Tabular - Consolidated	0.135** (0.067)	0.135* (0.070)	0.171* (0.092)	0.177** (0.089)
Observations	129	129	129	129
R^2 or pseudo R^2	0.040	0.052	0.052	0.071

Notes: The dependent variable is 1 if the dominated card is chosen. Robust standard errors are reported in parenthesis, and the probit specifications report marginal effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The regression estimates presented in Table 3 formalize the results concerning the likelihood of choosing the dominated card. The first two columns present linear probability models, in which the dependent variable is equal to 1 if the participant chose the dominated card. The coefficients on the Tabular and Consolidated indicators confirm the graphical evidence of Figure 6: the Tabular presentation is associated with a small

but insignificant reduction in the likelihood of choosing a dominated option, whereas the Consolidated treatment has a strong negative impact on the likelihood of selecting the dominated choice and the effect is significant at the less than 5% level.

Additionally, as can be seen in Column (1), for instance, the difference between the Tabular treatment and the Consolidated treatment is significant at the 5% level, providing clear evidence of attribute overload, with the reduction in attribute space for otherwise equivalent choices shrinking the fraction of people choosing a welfare-inferior GPR/prepaid card by about 14 percentage points. Given the choice treatments are balanced, it is no surprise that including our set of controls in Column (2) has little impact on these estimates. The same is true of using probit estimates (Column 3 and 4). Considering our estimates of the effects of the observables, increasing the stakes, one's income, one's cognitive acuity and financial literacy all reduce the likelihood that the dominated card will be chosen, but the effects, though appreciable, are not significant.

III.3 Best Card Choice

We have thus far seen that the Consolidated treatment significantly reduces the time that participants require to make a selection and also significantly reduces their likelihood of choosing a dominated option. The question remains whether the choice architecture treatments also influence the participants' ability to choose the *Best* card, given their preferences, and using our estimates of their discount factor and utility curvature. Regarding this question, our results are less straightforward. In Figure 6 we see that both reducing the number of attributes and using a table to summarize the attribute differences does increase the number of people who choose the Best card, but, not significantly so. Comparing the magnitudes, contrary to our worst card results, here the tabular presentation of the card details has the stronger effect on choice. Table 4 presents the same regression models as Table 3, except that the dependent variable is now equal to 1 if the

participant chose the Best card. While the coefficient on Tabular is large, indicating that the participants are 15 percentage points more likely to choose the best card when the many attributes are presented in tabular rather than textual form, as noted, the result is not quite significant at the 10% level ($p = 0.13$, to be precise). In fact, the only variable that has a significant impact on the likelihood of choosing the best card is the size of the Stakes, indicating that as the best option becomes more attractive relative to the other cards, participants are significantly more likely to choose it.

Table 4: Choosing the Best Card

	(1)	(2)	(3)	(4)
	OLS	OLS	Probit	Probit
Tabular	0.154 (0.103)	0.161 (0.101)	0.152 (0.100)	0.151 (0.096)
Consolidated	0.043 (0.107)	0.068 (0.103)	0.044 (0.109)	0.072 (0.104)
Stakes (in EU)		0.041*** (0.012)		0.043*** (0.016)
High Income (I)		-0.005 (0.088)		-0.011 (0.085)
High CRT (I)		-0.062 (0.086)		-0.053 (0.084)
High Fin. Lit. (I)		0.050 (0.089)		0.054 (0.086)
Constant	0.326*** (0.072)	0.240** (0.096)		
Tabular - Consolidated	0.111 (0.108)	0.093 (0.107)	0.108 (0.103)	0.079 (0.100)
Observations	129	129	129	129
R^2 or pseudo R^2	0.019	0.096	0.014	0.075

Notes: The dependent variable is 1 if the best card is chosen. Robust standard errors are reported in parenthesis and the probit specifications report marginal effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

III.4 Welfare

Given our estimates of the utility parameters for each participant, we are able to evaluate the welfare consequences of the choice interventions we have implemented. We know that consolidating the attributes significantly reduces the likelihood that the consumer will choose the dominated card and that presenting the fee information in tabular form has a substantial (though noisy and not quite statistically significant) effect on whether or not the consumer picks the best card. While this all may be true, are the welfare differences of any consequence?

To assess the welfare of our participants and to compare across them, we translate utility gains into dollar equivalents. Specifically, using the convex time budget utility parameter estimates that we used to create the cards for each individual, we first assess two differences: $EU(chosen) - EU(dominated) = u_1$ and $EU(best) - EU(dominated) = u_2$. Inverting these expected utility differences results in their dollar equivalents, $U^{-1}(u_1) = \$x_1$ and $U^{-1}(u_2) = \$x_2$. We then form the ratio of these two dollar amounts (x_1/x_2) to determine the fraction of the feasible utility gains (measured in dollars) that the participant achieved.

In the left panel of Figure 7 we present the mean fraction of the maximum dollar gains that are achieved. As the reader can see, there are differences by treatment but there is also considerable variation. As anticipated, both the Tabular and Consolidated treatments improve welfare and the differences are not small, approximately a tenth of the potential gains, but the confidence intervals are large too and therefore the improvements are not significant at conventional levels.

We mostly confirm these results in the first two columns of Table 5. In Column (1), we see that the Tabular treatment's effect on the chances of choosing the best card has a slightly larger effect on welfare than the Consolidated treatment's effect on preventing the worst card from being chosen. However, when we control for what is at stake (also

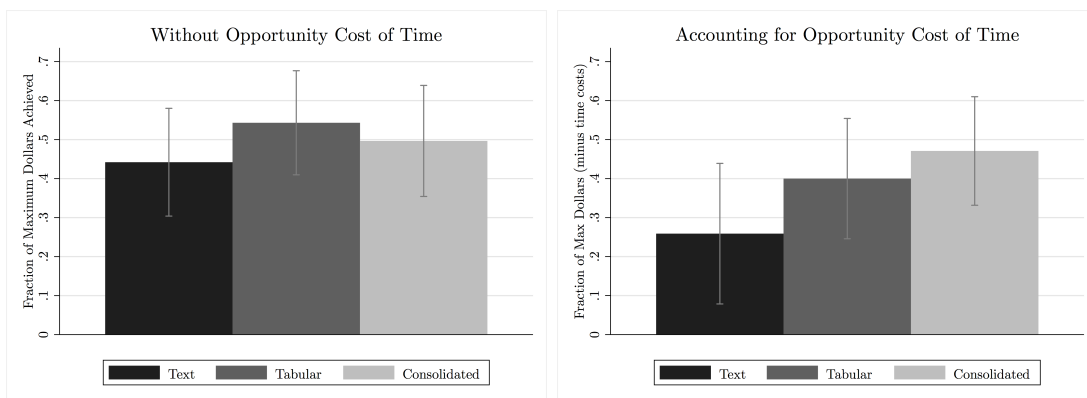


Figure 7: The fraction of the maximum dollars achieved (and the 95% C.I.s).

	(1)	(2)	(3)	(4)
	OLS	OLS	OLS	OLS
Tabular	0.101 (0.097)	0.105 (0.098)	0.141 (0.120)	0.155 (0.115)
Consolidated	0.055 (0.100)	0.071 (0.101)	0.212* (0.115)	0.229** (0.112)
Stakes (in dollars)		0.024*** (0.008)		0.037*** (0.010)
High Income (I)		-0.001 (0.082)		0.125 (0.095)
High CRT (I)		-0.009 (0.082)		-0.019 (0.091)
High Fin. Lit. (I)		0.035 (0.084)		-0.001 (0.091)
Constant	0.442*** (0.070)	0.370*** (0.099)	0.259*** (0.091)	0.103 (0.149)
Time cost deducted	No	No	Yes	Yes
Observations	129	129	129	129
R^2	0.009	0.047	0.026	0.099

Notes: The dependent variable is the fraction of the dollar gains available that are achieved; robust standard errors reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

in terms of dollars) in column (2), the treatment effects become more similar in size. Again, however, in both cases the welfare “bumps” that participants get in expectation from our interventions are noticeable but not significant. In addition, as was seen in our assessment of choosing the best card, stakes matter in the allocation of welfare too. In Column (2) when the stakes are larger, so too is welfare. In the end, worrying about choice architecture does appear to benefit the consumers in our experiment but our estimates on the size of this welfare gain lack precision.

One obvious problem with our welfare calculations so far is that we are not accounting for the opportunity cost of time. If, indeed, “time is money,” then we should account for the fact that some participants spend considerably more time struggling to make a card choice (or, relatedly, account for the possibility of “thinking aversion,” a la Ortoleva, 2013). To take into consideration the time cost of choice, in the right side of Figure 7 we subtracted the opportunity cost of the time spent choosing beyond that of the median participant. Specifically, we first calculated the median number of seconds required to choose (93, in our case) and then calculated the difference between each individual’s time and the median. We multiplied positive differences by our measure of the opportunity cost of time (the hourly average wage gain in the experiment), and then subtracted this from the numerator of the participant’s welfare ratio described above. When the difference was negative, the ratio was not “corrected.”²²

The shape of the bar chart on the right side of Figure 7 is as expected knowing what we do about how long it took to make choices (i.e., recall Figure 5). Because it took considerably longer to choose in the Text treatment than in the Consolidated treatment, the welfare gap between the two widens substantially.

After accounting for the opportunity cost of the time spent choosing in the second

²²Our principal motivation for the introduction of what amounts to a kink in the opportunity cost of time schedule is a desire to focus on the welfare losses of those who found the choice of cards an especially difficult one. As a practical matter, we also wanted to limit the influence of those who took almost no time at all, that is, those who did want to make an informed choice.

two columns of Table 5, we see that the welfare gains of our interventions grow and in the Consolidated treatment case the difference becomes significant. At the same time, while people chose quicker in the Tabular treatment too, and this does cause our estimate of the associated welfare increment to increase, it does so modestly. That said, though the 14 percentage point difference seen in Column (3) is not significant, it is economically relevant. Considering our other intervention, recall that it was the Consolidated treatment wherein participants saved the most time. This is reflected in the substantial increase in our estimate of how much better off participants are in this treatment compared to the baseline. Consolidated participants are now estimated to be more than a fifth better off than the Text participants and the difference is significant at the 10% level without controls. Lastly, this Consolidated treatment effect on welfare increases to almost a quarter and becomes significant at the 5% level when we, again, control for the other factors in Column (4).²³

III.5 Robustness

Our expansive experimental design allows us to examine the robustness of our results on a number of dimensions. To begin, because we had participants do both the convex time budget preference elicitation and the multiple price list elicitation, we can analyze our choice data re-ordering the cards ex-post based on the multiple price list utility parameters instead of those from the CTB. The comparison between the CTB estimates and the calculated parameter values based on the DMPL data are presented in Figures A4 and A5 in the Appendix. The DMPL estimates are calculated using the same methods as Andreoni and Sprenger (2012) and the comparisons across the two elicitation are similar to their findings. While the DMPL produces greater variation in participants' utility curvature, the upshot is that for only two people is there a difference in the utility

²³Of course, the welfare effect of the opportunity cost of the time spent choosing depends directly on how expensive it is to wait. We consider this more carefully in the next section.

ordering of the cards presented and therefore there is essentially no change in Figure 6. In other words, our results are not dependent on the preference elicitation protocol that we employ.

Given the demographics and behavioral information that we collected from our participants, we can also examine the possibility of heterogeneous treatment effects. This evidence is compiled in Tables A1 through A4 in the appendix. In each case, we split the sample at the median and examine whether participants who have low values for the stakes involved, low family income, low scores on the cognitive reflection task, or low financial literacy scores exhibit different responses to our interventions than subjects with high values for these characteristics. In each table, the first two columns describe the regression results for picking the worst card, the second two columns reports results on picking the best card and the remaining two columns show the results of our welfare analysis, accounting for the opportunity cost of time spent choosing a card.

In Table A1, we identify two effects. First, the Consolidated treatment has a much larger impact on not choosing the worst card for people with a lot (of dollars) at stake compared to other participants with not a lot at stake. Second, people without a lot at stake have substantially greater welfare gains from the Consolidated intervention than those with a lot to lose, mostly because the intervention allows them to make their choice quickly.

Considering differences in familial income (Table A2), our results become more pronounced relative to the Text/fine print baseline among lower income subjects. Within this population, far fewer chose the worst card while the point estimate for the more affluent participants is very close to zero. As a result, the low income participants benefited substantially from the Consolidation intervention, achieving a welfare bump more than a third of the possible gains.

The effect of the Consolidated treatment by scores on the cognitive reflection task

(CRT) can be seen in Table A3. In this case, high scorers on the CRT are more than 25 percentage points less likely to pick the worst card when in the Consolidated treatment than when in the baseline Text treatment, with a similar and slightly smaller reduction in worst card choice for the Consolidated treatment in comparison to the Tabular treatment (both effects being significant). For low scorers, however, there is no significant difference among treatments. This finding, interesting in its own right, allows us to assess a possible interpretation of the underlying mental processes behind the attribute overload phenomena that we document. One interpretation of the treatment effects is that they may materialize because the challenge of consolidating numerical attributes oneself can be overcome with the assistance of some of the choice architectures we provide. This can be seen as attribute overload by another name, or, a more specific mental pathway that constitutes one instantiation of a broader attribute overload phenomena. If the reduction in worst cards for the Consolidated treatment is coming about because the treatment is essentially reducing arithmetic difficulty, we would expect more pronounced treatment effects for those with more limited baseline numerical sophistication. In our case, this means we would expect the greater treatment effect for those with lower CRT scores, as the CRT has also been shown to be a measure of both reflexivity and basic numeracy (Szasi, Szollosi, Palfi, and Aczel, 2017). However, as shown, it is not the subjects with low CRT (and thus lower arithmetic numeracy) who are more susceptible to the treatment effects, but rather the reverse, suggesting the attribute overload we document is not coming about solely because of the difficulty in numerical processing alone, and, that something more general about the complexity of the Tabular treatment is thwarting good decision making.

Table A3 also shows that, even for CRT high scorers, the treatment effect does not seem to clearly translate into a significant increase in welfare (though there is marginal significance in this result). Lastly, in Table A4, we see that the Consolidated treatment helps both those who score low and those who score high in financial literacy to avoid

the worst card and achieve higher welfare. However, the effect is only significant for the low financial literacy group.

Considering the robustness of our welfare estimates, we examine the implications of our utility specification, the discount rate data we employ, and, our assumptions about the opportunity cost of our participants' time. To begin, as mentioned above, by construction, our discount factor estimates (because they are daily) are bound to be close to 1. In addition, we observe that the mean of our utility curvature estimates is also close to 1. What if we simply assume all our participants are money maximizers (i.e., $\alpha = \delta = 1$)? How would this affect our welfare results? In Table A5, we reproduce Table 5 with just one change - now the dependent variable is the dollar equivalent of the utility gains assuming simple additive utility, $U(c_1, c_{1+k}) = (c_1) + (c_{1+k})$. As one can see, proxying the preferences with money maximization, leads to slightly smaller welfare gains for the Consolidated and Tabular treatments because we throw away some important variation in preference parameters, but the results are not that much worse because the preferences are close to $\alpha = \delta = 1$ by design.

Focusing further on our time preference data, as part of our financial literacy survey in the first part of the experiment, we asked participants about their credit cards - specifically, the interest rates on these cards. In Table A6, we examine what happens if we substitute the discount factors implied by these credit card interest rates for the deltas we collected during the experiment, the underlying assumption being that participants who are not liquidity-constrained (like students with large family incomes) should discount rewards at the interest rate available to them. This alternative performs poorly. Not only do the treatment differences vanish, the stakes involved fail to predict the gains achieved, and, overall, the models fit poorly. Given that we know the Consolidated treatment leads to significantly fewer choices of a dominated product, the fact that no treatment difference in the welfare results for table A6 materializes makes us skeptical of the appropriateness of this alternative discount factor (or, at least, of participants

correctly reporting their credit card interest rates).

Lastly, for our baseline analysis of welfare we set the opportunity cost of time at the stakes involved in the experiment, roughly \$36 per hour. One might think that this rate seems high and that the rate should be closer to that achievable at a job on campus. Alternatively, one could argue that many students reject campus jobs at the going rate because the wage is below their opportunity cost of time. Further, another plausible estimate of time cost is one's expectation of the wage achievable upon graduation. Because all these estimates seem reasonable, in Table A7, we offer welfare estimates based on each of them. It will be no surprise that the higher the opportunity cost, the larger the welfare gains accruing to the two information architecture treatments (the Consolidated treatment, in particular), but we find that these gains remain positive and significant to the cutoff value of \$22.50 per hour.²⁴

III.6 External Validity

In this subsection we explore the relationship between our experimental findings and naturally occurring products in the prepaid card market. In particular, we consider a set of prepaid cards classified along the dimensions of consumer welfare and choice architecture, and explore the consequences of the CFPB regulatory change described earlier. The results, while imprecise, are consistent with what we observed in the lab.

Outside the lab, of course, reliance on proxies for card value and choice architecture is almost inevitable. We consider two card value proxies: a measure determined by consumer advocates and experts in the prepaid card market and, perhaps more convincingly, a record of complaints made to the CFPB regarding specific cards, under the assumption that complaints track consumer (dis)satisfaction and welfare. The first of these

²⁴This threshold exceeds the wage for *all* college graduates, which is closer to \$20 per hour, but reflects the particular labor market prospects of our participants who are on the right tail of the income distribution. At the lower rate, the Consolidated treatment effect is of course smaller (0.159) and not quite significant ($p = 0.117$) at the 10% level.

comes from a 2016 Consumer Reports review of 20 prepaid cards, and is derived from the application of a simulated user’s spending patterns to a card’s actual fee structure, as summarized in that review.

Proxies for choice architecture are, if anything, more elusive: CFPB estimates that there are thousands of differentiated products and, to date, there is no exhaustive assessment or classification of the presentation of fees and card attributes.²⁵ For a small subset of these (less than two dozen), however, the same consumer advocates have provided such a measure: the aforementioned Consumer Reports review rated the cards on three other dimensions, one of which was “fee accessibility and clarity,” measuring “the ease of finding and understanding information and disclosures about the fees.” There were (just) two categories, which we now denote “lower clarity” and “higher clarity.” Further, given the advocacy arm of Consumer Reports’ prior advocacy for the specific CFPB reforms, we believe that “lower clarity” cards resemble our baseline Text treatment, while “higher clarity” cards are closer to our Tabular Treatment. Figure 8 presents the distribution of card values by clarity, and suggests that when fee structure is less transparent, fees for the representative user are higher or, in other words, card value is lower: almost half of all the “more clarity” cards are deemed to be either “very good” or “excellent” value, but none of the “less clarity” cards are.

Obviously, Figure 8 reflects only simulated user behavior rather than actual consumer behavior, a limitation of the Consumer Reports data alone. However, we can turn to an alternative data source, one that reflects the behavior of actual consumers: a CFPB complaint database.²⁶ The database is a collection of complaints on a range of consumer financial products and services, including GPRs. For our purposes, we use the com-

²⁵By one recent count, just over 2000 prepaid products were known to exist, with roughly one half of these having unique names, and, the others being different versions of these distinguished by different terms and conditions.

²⁶As is true in the rest of paper, the research, methodology, analysis, and conclusions expressed in this section are those of the authors only. They do not reflect or represent the complaint reporting methodology, analyses, or observations of the Consumer Financial Protection Bureau.

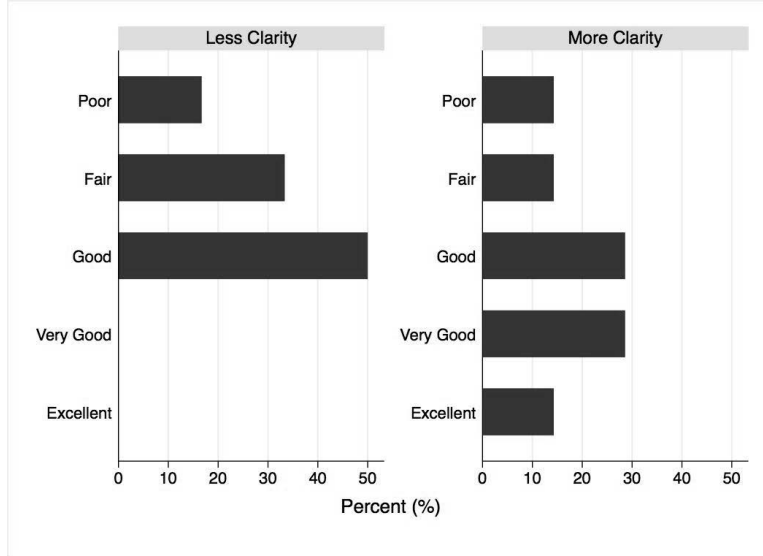


Figure 8: Histogram of Card Value Measure by Consumer Report Clarity Score

plaints about specific GPRs to proxy for product-specific consumer welfare. Given the same Consumer Reports clarity measure, we are able to provide some limited evidence of the CFPB regulatory reform that went into effect on April 1, 2019.

The reform, as discussed in the Introduction, required prepaid card issuers to improve the transparency of their fee and terms disclosure in a standardized format akin to our Tabular treatment. Our experimental findings suggest a particular pattern of complaints over the pre and post period. For cards already in the higher clarity group, the reform presumably changed little, and so constitute a control group of sorts. The trends in complaint counts for this subset of cards is therefore the baseline against which to measure other non-policy related changes to monthly card complaints. For lower clarity cards, we expect the reform to bind and for consumer welfare to increase. A comparison of the differences in complaints between the two clarity groups pre- and post-reform - that is, the usual difference-in-difference - then provides an estimate of the treatment. Given the small number of cards reviewed in the Consumers Report - and therefore the even smaller numbers of control and treated units - the standard errors are of course too large to produce precise estimates, but the raw pattern is consistent with our experimentally-

informed predictions.

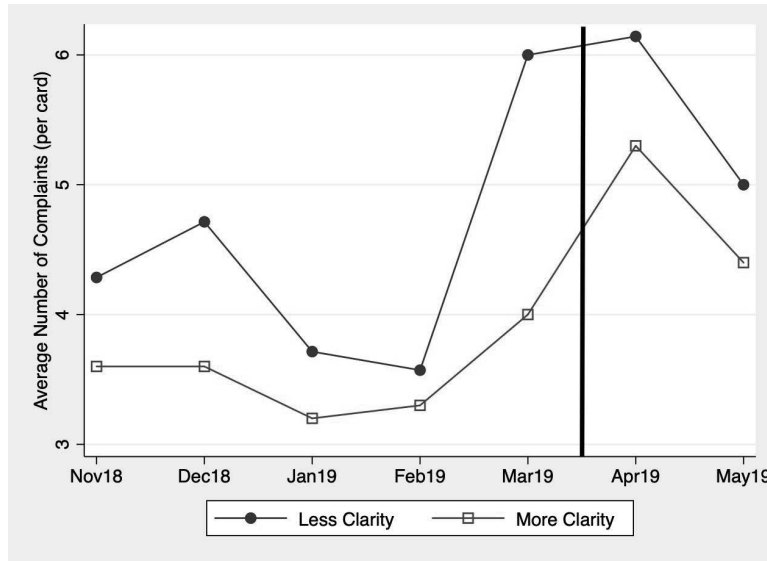


Figure 9: Complaints to CFPB Pre and Post Policy Reform (by Pre-Reform Clarity Type)

Figure 9 illustrates the difference in average reported complaints (per card) between the two types of cards. As one can see, the number of complaints associated with both the lower clarity and higher clarity cards is relatively stable in the months prior to the intervention, with the less clear cards receiving more complaints on a consistent basis through the end of March 2019, as expected. In April, however, the complaint difference begins to shrink and shrinks further in May. In fact, the difference in this difference from March to April (the immediate scope for the regulation’s onset) falls by approximately half. Alternatively (and more conservatively), pooling each of these pre and post periods to generate a pre-period monthly average and a post-period monthly average, the equivalent difference in the difference falls by closer to 15%. To reprise an earlier observation, our comparisons are based on a very small number of cards, and we are therefore reluctant to make too much of them. The point, however, is that using the limited data available from the field appears to corroborate the lab result, namely, that changing choice architecture to a more systematic tabular-style presentation can improve consumer well-being.

IV Conclusion

The purpose of this study is to evaluate policy interventions into the prepaid card market that are based on the principles of choice architecture. Specifically, we designed experimental treatments that aim to make choices more comparable and the presentation of attributes more parsimonious. Given the prepaid card market has expanded to \$300 billion, the implications of choice and welfare enhancing policies are surely large.

The unique elements of our experiment allow us to estimate consumer-level utility functions that allow us to both individualize the set of prepaid cards that consumers face and conduct a welfare analysis in addition to a more precise choice analysis. We find that, when placed in conditions that simulate some products in the status quo marketplace, consumers mostly make poor choices (based on their own estimated preferences). When we examine a disclosure in the spirit of the CFPB's short form, one that aims to present information clearly and concisely, in a tabular format, and in a way that facilitates comparing terms across products, we find that it does help - more participants pick the card that should maximize their stated preferences. When we consider an even more parsimonious choice architecture designed to reduce attribute overload too, we find that the number of consumers making the best choice improves only slightly (relative to the status quo) but that very few consumers make dominated choices (relative to either alternative choice architecture), with this later finding consistently statistically significant at conventional levels. This enhanced architecture also dramatically reduces the time required to make a choice. Bearing this in mind, based on the preferences of our participants and a measure of the opportunity cost of time, the initial intervention improves consumer welfare somewhat but the enhanced policy improves welfare substantially. Furthermore, considering heterogeneous treatment effects, the enhanced architecture interacts with many of the important characteristics of our participants. This architecture appears to differentially help those with a lot at stake, low family income, high cognitive ability

and low financial literacy, in particular, to not make the worst choice. Given the demographics of the broader industry in which these cards are now common, these seem like exactly the people who would benefit most from a policy intervention. Further, we find tentative support for our results in the field, based on the relationship between, on the one hand, fee clarity and, on the other, card value and card complaints for a subset of prepaid cards.

Finally, one strength of the experimental design is that while the total amount of information varied across treatments, the utility-relevant information was held constant. This allowed for clear, causal inference about the effects of information and presentation on choice. However, as mentioned above, applying this result directly to the field would be challenging since the structures of most prepaid products do not allow one to alter the total amount of information without also altering the utility-relevant information. Indeed, it is important to consider whether optimal parsimony of product disclosure is possible given the complexity of the products in the marketplace. Pew's survey of prepaid cards (2012), for example, found the most fee-laden prepaid cards have more than 20 fees, and this does not consider the many terms and conditions that accompany such products. While some reduction in the attribute space is possible, it is likely that in many contexts the parsimony found in our enhanced choice architecture will be hard to achieve. In these contexts where consolidation is either difficult or misleading, regulators will be challenged to realize the full extent of the benefits of consolidated disclosure. In this case, regulation of the financial products themselves, and not just the disclosures on the packaging, may be in order. From the choice architecture perspective, when financial products are complex and have a large number of attributes, the impact of improving the organization and disclosure of those attributes may not be enough on its own. Large improvements in choice and consumer welfare may require a more elemental reduction in the number of attributes.

V References

Agnew, J. R., Szykman, 2005. Asset allocation and information overload: The influence of information display, asset choice and investor experience, *Journal of Behavioral Finance* 6(2), 57-70.

Anderson, S., Harrison, G., Lau, M., Rutstrom, E., 2008. Lost in State Space: Are Preferences Stable? *International Economic Review* 49, 1091-1112.

Andreoni, J., Sprenger, C., 2012. Estimating Time Preferences from Convex Budgets. *American Economic Review* 102, 3333-3356.

Andreoni, J., Callen, M., Hussain, K., Khan, M.Y., Sprenger, C., 2016. Using Preference Estimates to Customize Incentives: An Application to Polio Vaccination Drives in Pakistan. NBER Working Paper No. 22019.

Augenblick, N., Niederle, M., Sprenger, C., 2015. Working over time: dynamic inconsistency in real effort tasks. *Quarterly Journal of Economics* 130, 1067-1115.

Bertrand, M., Morse, A. 2011. Information Disclosure, Cognitive Biases, and Pay-day Borrowing. *Journal of Finance* 66(6): 1865-1893.

Besedes, T., Deck, C., Sarangi, S., Shor, M. 2012. Decision-Making Strategies and Performance Among Seniors. *Journal of Economic Behavior and Organization* 81, 524-533.

Besedes, T., Deck, C., Sarangi, S., Shor, M. 2015. Reducing Choice Overload Without Reducing Choices. *Review of Economics and Statistics* 97(4), 793-802.

Carpenter, J., Huet-Vaughn, E., 2017. Real Effort Tasks, forthcoming in: Schram, A., Ule, A. (Eds.), *Handbook of Research Methods and Applications in Experimental Economics*. Edward Elgar, Cheltenham, UK.

Choi, J.J., Laibson, D., Madrian, B.C. 2010. Why Does the Law of One Price Fail? An Experiment on Index Mutual Funds. *Review of Financial Studies* 23(4): 1405-1432.

Cole, A., Greene, C., 2016. Financial inclusion and consumer payment choice. Federal Reserve Bank of Boston Working Paper No. 16-5.

Consumer Financial Protection Bureau, 2014a. Study of prepaid account agreements. CFPB Working Paper.

Consumer Financial Protection Bureau, 2014b. CFPB proposes strong federal protections for prepaid products, CFPB Newsroom.

Consumer Reports, 2017. Prepaid card buying guide: getting started, Consumer Reports Buying Guides.

Consumer Reports, 2016. Prepaid cards: How They Rate 2016.

Diehl, K., Kornish, L.J., Lynch, J.J.G., 2003. Smart Agents: When Lower Search Costs for Quality Information Increase Price Sensitivity. *Journal of Consumer Research* 30, 56-71.

Fasolo, B., McClelland, G., Todd, P., 2007. Escaping the tyranny of choice: when fewer attributes make choice easier. *Marketing Theory* 7, 13-26.

Fischbacher, U., 2007. z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics* 10, 171-178.

Frederick, S., 2005. Cognitive Reflection and Decision Making. *Journal of Economic Perspectives* 19, 25-42.

Greene, C., Schuh, S., Stavins, J., 2016. The 2014 Survey of Consumer Payment Choice: Summary Results, Federal Reserve Bank of Boston working paper 16-3.

Greene, C., Shy, O., 2015. How are U.S. consumers using general purpose reloadable prepaid cards? Are they being used as substitutes for checking accounts?, Federal Reserve Bank of Boston Working Paper No. 15-3.

Hayashi, F., Cuddy, E., 2014. General purpose reloadable prepaid cards: penetration, use, fees, and fraud risks, Federal Reserve Bank of Kansas City Working Paper

RWP 14-01.

Hayashi, F., Cuddy, E., 2015. Recurrent overdrafts: a deliberate decision by some prepaid cardholders?, The Federal Reserve Bank of Kansas City Working Paper RWP 14-08.

Holt, C., Laury, S., 2002. Risk Aversion and Incentive Effects. *American Economic Review* 92, 1644-1655.

ICF International, 2014. "Summary of Findings: Design and Testing of Prepaid Card Fee Disclosures," (http://files.consumerfinance.gov/f/201411_cfpb_summary-findings-design-testing-prepaid-card-disclosures.pdf).

ICF International, 2015. "Final Report of Findings: Post-Proposal Testing of Prepaid Card Disclosures," (http://files.consumerfinance.gov/f/201510_cfpb_report-findings-testing-prepaid-card-disclosures.pdf).

Iyengar, S. S., Lepper, M. R., 2000. When choice is demotivating: can one desire too much of a good thing? *Journal of Personality and Social Psychology*, 79(6), 995-1006.

Johnson, E., Shu, S., Dellaert, B., Fox, C., Goldstein, D., Haubl, G., Larrick, R., Payne, J., Peters, E., Schkade, D., Wansink, B., Weber, E., 2012. Beyond nudges: tools of a choice architecture. *Marketing Letters* 23, 487-504.

Knoll, M.A.Z., Houts, C.R., 2012. The Financial Knowledge Scale: An Application of Item Response Theory to the Assessment of Financial Literacy. *Journal of Consumer Affairs* 46, 381-410.

Lee, B-K., Lee, W-N. 2004. The effect of information overload on consumer choice quality in an on-line environment. *Psychology and Marketing*, 21(3), 159-183.

Meier, S., Sprenger, C., 2015. Temporal stability of time preferences. *Review of Economic Studies* 97, 273-286.

Ortoleva, P., 2013. The Price of Flexibility: Towards a Theory of Thinking Aversion. *Journal of Economic Theory* 148, 903-934.

Peters, E., Dieckmann, N., Vastfjall, D., Mertz, C., Slovic, P., Hibbard, J., 2009. Bringing meaning to numbers: the impact of evaluative categories on decisions. *Journal of Experimental Psychology. Applied* 15, 213-227.

Pew Charitable Trusts, 2012. Loaded with uncertainty: are prepaid cards a smart alternative to checking accounts? *Consumer Financial Security*.

Pew Charitable Trusts, 2014. Why Americans use prepaid cards: a survey of cardholders' motivations and views.

Ratcliffe, C., Congdon, W.J., McKernan, S.-M., 2017. Prepaid Cards at Tax Time and Beyond: Findings and Lessons from the MyAccountCard Pilot. *Journal of Consumer Affairs*, forthcoming.

Rhine, S., Jacob, K., Osaki, Y., Tescher, J., 2007. Cardholder use of general spending prepaid cards: a closer look at the market, *The Federal Reserve Bank of Chicago Working Paper*.

Samek, A., Hur, I., Kim, S-H., Yi, J-S. 2016. An experimental study of the decision process with interactive technology, *Journal of Economic Behavior and Organization* 130, 20-32.

Scheibehenne, B., Greifender R., Todd P.M., 2010. Can There Ever Be too Many Options? A Meta-Analytic Review of Choice Overload. *Journal of Consumer Research* 37, 409-425.

Schildberg-Hörisch, H., 2018. Are risk preferences stable? *Journal of Economic Perspectives* 32, 135-154.

Schwartz, B., 2004. *The paradox of choice*. Harper Perennial, New York, NY.

Sloane, T., 2017. 14th Annual U.S. Open-Loop Prepaid Cards Market Forecast. 2017-2020, Mercatur Advisory Group.

Smith, V., 1994. Economics in the Laboratory. *Journal of Economic Perspectives* 8, 113-131.

Soll, J. B., Keeney, R. L., Larrick, R. P. Consumer misunderstanding of credit card use, payments, and debt: causes and solutions, *Journal of Public Policy and Marketing* 32(1), 66-81.

Szaszi, B., Szollosi, A., Plafi, B., Aczel, B., 2017. The cognitive reflection test revisited: exploring the ways individuals solve the test, *Thinking & Reasoning*, 23:3, 207-234.

Thaler, R., Sunstein, C.R., 2008. *Nudge: improving decisions about health, wealth, and happiness*. Yale University Press, New Haven, CT.

Thaler, R., Sunstein, C.R., Balz, J., 2012. Choice Architecture, *The Behavioral Foundations of Public Policy*, pp. 428-439.