U.S. Securities and Exchange Commission

Office of the Investor Advocate

OIAD Working Paper 2022-03

June 2022

Choice Set Complexity

Tradeoffs and Decision Strategies in an Investment Case Study

BRIAN SCHOLL, Chief Economist, Office of the Investor Advocate, Securities and Exchange Commission, 100 F St., NE, Washington, DC 20549, schollb@sec.gov, ORCID ID: 0000-0001-5088-6952

ABSTRACT

I study the complexity of evaluating tradeoffs between features of choice set options by experimentally assigning a gradient of menus (792 total conditions) of varying complexity levels and observing subjects' choices. While this situation exists in a variety of settings, I focus on a case study in a financial investment setting where: payouts are inherently in a numeraire good; I can construct a choice set for which the strictly dominant strategy is unambiguous; and I can randomly assign choice sets to vary tradeoffs. In this setting, I am able to experimentally estimate the loss associated with the gradient-increase of complexity without specific preference assumptions. I conservatively estimate an increase from low to high complexity increases menu-normalized fees by up to 20 percentage points, a projected welfare loss of 60 percent of subjects' initial investment. There is little evidence that effects moderate by subgroup. Subjects pursued naive and highly unstable decision strategies in their choices. Salience provides a complementary, not competing mechanism of choice, while extremely high sophistication and a cost-quality empirical heuristic demonstrate the best chance of moderating the complexity effect.

ACKNOWLEDGEMENTS

This paper benefited from contributions from and discussions with Brigitte Madrian, Rick Larrick, John Lynch, Dan Silverman, Rick Fleming, Alycia Chin, Abhi Jain, Angela Fontes, Mark Lush, Maggie Lehr, Seth Cohen, Jonathan Cook, and Steven Nash.

---

1 Introduction

Household decisions in modern economies are often made amid a sea of complexity. Large varieties of features and options all contribute to a rich choice environment that potentially allows the individual or household to curate an experience tailored to their preferences - for example, nut- and gluten-free granola, or cars with different interior designs, trunk space, trim and other features. More choice features potentially enable the individual more options to identify products that suit their individual desires. The dominant view in economics is that more feature options are an unambiguous good: by trading off features in the context of budgetary and other constraints, agents maximize utility subject to those constraints. In the canonical approach, one does this with full information, full knowledge and costless decision-making, observed selections enable the researcher to reveal an underlying set of fixed preferences.

For real people in a real world setting, the decision-making process can be overwhelming and costly and clunky and murky, so that many frictions have been introduced that inhibit the optimization approach of the canonical model (see, for example, DellaVigna (2009), Harstad & Selten (2013), Rabin (2013), Golman et al. (2017)). In this paper, I use a highly sophisticated continuous-assignment experiment to provide evidence that contributes to a recent literature on the adverse consequences of various forms of complexity (e.g. Oprea (2020), Sethi-Iyengar et al. (2004), Chin et al. (2021), Scholl et al. (2020*b*)) by introducing a new concept of complexity: **choice set complexity**, which refers to *the complexity generated by implied tradeoffs between features of items within the choice set.* This phenomenon is a menu effect that has hitherto not been pursued in the economic literature (see DellaVigna (2009) for a review), with other recent literature regarding complexity of tasks, information, or language having looked at very different complexity concepts. Most closely related to the current paper, Iyengar & Lepper (2000), Sethi-Iyengar et al. (2004), Carvalho & Silverman (2019) and others explore the effects of increases in the size of the choice set; in my setting, I fix the size of the choice set and introduce increasingly complex tradeoffs between features of the items on the menu in order to examine decisionmaking quality. With a careful curation of choice-set conditions, my experimental framework al- lows me to examine a very refined gradient of change in choice set complexity using nearly 800 different experimental assignments.

To fix ideas about choice set complexity, consider the two menus of Figure 1, each with three products (items A,B,C, and items D,B,E), and each product with two features to consider (Feature 1 and Feature 2). The numbers indicate each product's desirability rank in terms of the respective feature so that on the left menu, item A is best in terms of both feature 1 and feature 2. With these rank orderings, the menu on the left presents a low choice set complexity menu: a consumer forced to pick one choice item from the menu does not need to make a tradeoff because Choice A is most desirable in terms of both features. The menu on the right however, is somewhat different (even though item B is on both menus), the rank ordering of features is quite different. In this menu, Choice D is best in terms of Feature 1, but worst in terms of Feature 2, while Choice E is worst in terms of Feature 1 and best in terms of Feature 2. Thus, in this case the consumer must make a tradeoff between the features in making a choice, which could be challenging. Economic theory provides guidance on how such tradeoffs should be made, including activities re- quiring the individual to value and weigh the features. Yet, this task of optimization may be cognitively or emotionally burdensome, and may introduce the potential for calculation errors. In fact, some cases may simply be inevaluable to the individual - for example, when evaluating certain medical situations or treatments, or say when evaluating ethical situations such as the classic Trolley Dilemma in which the decision-maker chooses between the lives of individuals. In any event, clearly the menu on the right presents potential challenges that are not present in the menu on the left even if the extent and the consequences of such challenges varies situationally.

The choice set complexity phenomenon can arise in a variety of contexts, but can be difficult to study because preferences are typically identified using the observed choices individuals make: we use those selections to parameterize preferences, with the assumption that individuals landed at the optimal choice. In my experimental setting, there is a unique and unambiguous strictly dominant strategy for optimal choice, and my design eliminates the possibility that any other features of the choice set influence decisions. Indeed, my particular set-up enables me to develop a plausible estimate of welfare loss in the choice scenario without any assumptions on individual preferences; in fact, because my choice set

Choice	Feature 1	Feature 2
A	1	1
B	2	2
C	3	3

Choice	Feature 1	Feature 2
D	1	3
B	2	2
E	3	1

Figure 1: An example to illustrate a low vs. a high complexity menu. The menu on the left side is a low choice set complexity menu: Choice item A has the best rank preference in terms of both Feature 1 and Feature 2 (blue highlights), while Choice C has the worst rank on both features (red highlights). The menu on the right side is a high complexity menu: Choice D is best in terms of Feature 1 (blue), but worst in terms of Feature 2 (red), while Choice E is Best in terms of Feature 2 (blue) and worst in terms of Feature 1 (red). Choice B appears on both menus. Low complexity in this case is generated by a positive correlation in the preference ordering of both features (Choice A strictly dominates in all dimensions), while the menu on the right does not offer an option that strictly dominates in all features.

will involve payoffs in a numeraire good, I can directly observe welfare losses without an optimization model. I pursue this approach not because I reject the standard optimization approach (see Harstad & Selten (2013) and Rabin (2013) for some debate), but simply to avoid debate over a specific structural form of optimization and to minimize the number of assumptions I make in the context of evaluating choice decisions.

Specifically, to study the phenomenon of Choice Set Complexity, I pursue a case study using an incentivized purchasing task where subjects select one or more financial products from a choice set comprised of other similar products. In this context, each choice option is an S&P 500 index mutual fund; these are real funds, with identifying information anonymized so that any other preferences, brand affiliations and so forth are plausibly eliminated from consideration. One fascinating feature of the focus on such a choice set is that these products are essentially identical (gross returns, risk, strategy, etc.), but even in this financial sector context the market discipline mechanism has failed to eliminate price dispersion so that costs vary considerably (Hortaçsu & Syverson (2004)). Following other work that has focused on a similar choice set in other contexts (e.g. Choi et al. (2009), Fisch & Wilkinson-Ryan (2014)), subjects are presented with an apparent secondary feature of returns since inception, which is a function of the fund's start date rather than a measure of product quality. Because this choice consists of identical products offered with different costs, the strictly dominant strategy is unambiguously to put all money into the lowest cost fund regardless of the specific choice set the subject receives - and this strategy does not depend on any specific assumptions about individual preferences. To generate complexity, I curate a candidate set of twelve funds, randomly assigning five to each sub- jects' menu in a randomized order. By doing so, I randomly assign menus - a total of 792 combinations (12 choose 5), effectively meaning an almost unprecedented 792 experimental conditions - allowing me to provide a fine and continuous gradient of complexity and its effect on subjects' choices; in less complex choice sets, subjects receive a choice set in which the relationship between returns and cost involves no trade-off between the features (the cheapest option also ostensibly has the highest quality), while in the high complexity choice sets, subjects must give up a desirable property of one feature in order to gain a desirable property on another feature.

Results in this case strongly suggest that increasing the complexity of the choice set leads to a meaningful decrease in decision quality. An increase in complexity from the lowest levels to the highest results in an approximate 16 to 20 percentage point increase in menu-normalized fees, and there is little to no evidence that the effect moderates by sub- group. I pursue analysis of the experimental data using a highly conservative randomization inference framework (see Lin (2013), Athey & Imbens (2017), Freedman (2008), Freedman et al. (2008)), develop a placebo test and pursue other methods that suggest that my results are highly robust and not a product of design or estimation approaches. I also conducted a replication study with some of the original subjects: results were largely identical in both studies and for between-subjects and within-subjects estimation, with no evidence of learning between administrations for the repeat subjects. Because both the initial experiment and the replication study were fielded on a high-quality nationally representative probability sample survey with a total of nearly 7,000 subjects, my paper provides not only the customarily high internal validity, but also rarely encountered population-generalizable experimental findings.

To gain additional insight, I use descriptive analysis to pursue the decisioning mechanism. Cluster analysis suggests that participants pursued a set of very different - and highly unstable - decision strategies. The most common and most stable strategy was pure naive diversification; while clearly not an optimal strategy and although often vilified in the literature on portfolio choice (Benartzi & Thaler (2001), Choi et al. (2009)), naive diversification may actually be a reasonable strategy in the choice setting that I consider, particularly for low-sophistication individuals, as it results in a smaller welfare loss than some other strategies that were pursued. (Of course, this does not mean that naive diversification would necessarily be a benefit in all investment menu settings - even here it is clearly a welfare loss vs. the optimal strategy). A very small fraction of participants pursued the strictly dominant strategy, and for a group of subjects that repeated the experiment at a later date, few of these repeated the dominant strategy. There is some evidence that even relatively higher sophistication individuals pursue a weaker form of suboptimal diversification and may alternatively chase higher returns or low fees depending on the complexity of the choice set.

In other analysis, I examine alternative theories of decisionmaking including the salience models of Bordalo et al. (2013) (whether features' values stand out in the choice set) and the more common usage of the salience concept as for example in Hartzmark & Sussman (2019) (whether information is prominent); while there is evidence that these mechanisms may also be at work in the current setup, the complexity effect remains so that salience and complexity effects may be complementary. My evidence suggests that the mechanism though which complexity could be eliminated is with belief in a specific empirical fact about financial markets that has been established in the literature - namely that lower fees result in higher post-fee returns (rather than a "you get what you pay for" belief; a concept developed more fully in Scholl & VanEpps (2022)) - yet this empirical fact may not necessarily hold in other domains or may fall apart in a specifically curated choice set. A small group of subjects that scored in the top 6.5 percent of respondents on the mutual fund knowledge index of Scholl & Fontes (2022) also seem to have a reduced, albeit still present, complexity effect. Because trying to raise a broad cross section of investors to this level of sophistication seems elusive, these results suggest that a more promising avenue to minimize the complexity effect may be to target menu construction at the source rather than trying to arm individuals to avoid the complexity trap.

The consequence of the choice errors that subjects exhibit in my experimental setting are large, meaningful and real. Continuing with the investment choice case study, a simi- lar decision to my experimental context involves retirement plan choices by participating workers. Figure 2 provides the histogram of complexity using my choice set complexity measure computed for the full universe of actual retirement accounts: the distributional mass is concentrated in the high complexity range (negative values are increasingly lower complexity, positive values are increasingly higher complexity). My experimental setup closely approximates the decision environment that consumers face when making real-life retirement choices. In that light, estimates of the complexity effect within the experiment itself are extremely consequential: with assumptions similar to Benartzi & Thaler (2001) I project subjects' choices into a welfare loss of the 6,900 study participants in the provided choice set over a 25-year investment horizon. This projection yields an average nominal loss of USD 63,000 on an initial USD 100,000 investment - about USD 430 million for just those 6,911 task completions. In fact, this may be an extremely conservative estimate of the complexity effect because my experimental design so sharply eliminates other choice tradeoffs that would increase complexity in this and other settings and lead to further deviations from the optimal strategy. That is, I consider only two features to trade off, but often individuals will be faced with multiple features to compare. They also may need to weigh recommendations about products or other considerations, or be faced with more consequential magnitudes of feature tradeoffs than available in my particular choice set which itself represents relatively low feature dispersion (price dispersion) than is observed even in more general mutual fund choice problems.

This paper continues as follows: Section 2 describes the concept of choice set complexity in greater detail, and provides a topology of related literature on complexity in consumer and investment settings. Section 3 describes the experimental set-up and identification; Section 4 discusses experimental results, subgroup effects, and a placebo test for both the initial fielding and the replication study. Section 5 probes decision elements and

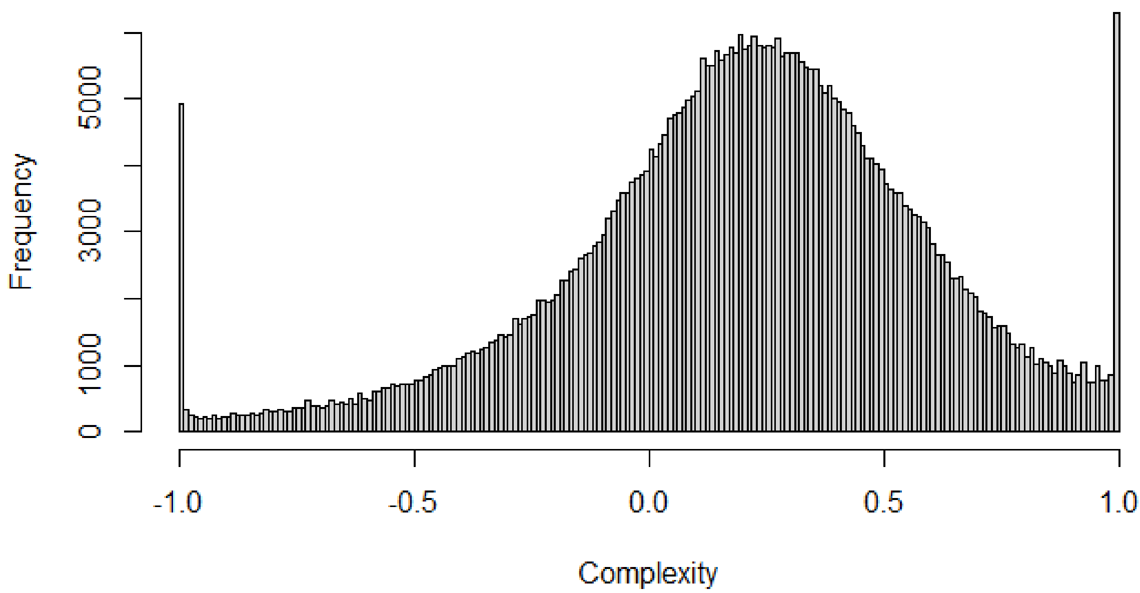


Figure 2: Histogram of Intra-Fund Correlations for retirement plans from Form 5500 fillings. Source Department of Labor and author's calculations.

decision strategies. Section 6 estimates the welfare implications of the study, and Section 7 concludes.

2 Choice Set Complexity Concepts and Literature

The dimension of complexity I study in this paper is choice set complexity, by which I mean the complexity of inter-attribute trade-off faced by an individual facing the decision problem. There are various forms of complexity that have been examined in the literature in the context of investor or consumer decision-making, and it is helpful to provide some topology here. One line of research involves the number of choice items that investors or consumers are presented with in choice problems. I refer to this form of complexity as **cardinality complexity** because it relates to the size of the choice set. Cardinality complexity has been examined rather extensively in consumer and investment choice problems, with important contributions by Iyengar & Lepper (2000), Carvalho & Silverman (2019), Sethi-Iyengar et al. (2004), Greifeneder et al. (2010), Besedeš et al. (2012), Brocas et al. (2015), Phatak (2012), Kempf & Ruenzi (2006) and meta-analysis by Scheibehenne et al. (2010). This work has largely demonstrated a curse of more choice. Carvalho & Silverman (2019) use experiments on constructed investment options to find that more options lead to poorer investment choices, particularly among less sophisticated individuals. Sethi-Iyengar et al. (2004) use 401(k) plan participation rates to study the effect of more choices on participation. They find that participation is higher in plans that offer only a handful of investment options. Besedeš et al. (2012) used choice tasks designed to be similar to choice tasks in the context of choosing investments in retirement savings, choosing health insurance plans, and selection of prescription drug insurance plans in settings where the number of menu items is large. Similar to the current study, their set-up allows for rank ordering of the quality of the options. They find that subjects were less likely to match with the optimal plan as the number of options increase; older subjects were particularly impaired, potentially relying on sub-optimal decision rules. Brocas et al. (2015) find similar impairment for older cohorts. Phatak (2012) also finds that the number of choices leads to deleterious effects on decisionmaking, although he suggests that oversimplification may also have adverse consequences. These applications are not the first to consider complexity arising from the number of choices - in psychology, marketing and consumer science, the effect of more choices on decision quality has been explored extensively (see, for example, the reviews in Bettman et al. (1998) and Bettman et al. (2008)).

Another dimension of complexity that may be termed **product complexity** relates to structure of assets, pricing or payment schemes. Some contributions to this line of research include the study of complex cost structures in Kalaycı & Serra-Garcia (2012), complexity in retirement fee structures in Tse et al. (2016), complexity in tax rules and structures in Abeler & Jäger (2013), complexity in cell phone tariff structures by Friesen & Earl (2015), and complexity in tax incentive schemes by Rees-Jones & Taubinsky (2016). The recent contribution by Oprea (2020) is related to this work; he pursues **task complexity** and individuals' willingness to pay to avoid implementing complex rules.

In other work, scholarship and policy efforts have often combined in **complexity abatement** studies. Innovations along these lines have often been gathered under the rubric of choice architecture, and have often led to operationalization of specific approaches designed to help make decisions easier for investors (see Thaler et al. (2013)). Beshears et al. (2013) provide a highly influential policy-relevant example employing an experiment to reduce complexity from a multi-stage decisioning problem, testing a simplification that reduced a complex decision problem into effectively a single participation decision.

Informational complexity, centers around the amount or organization of information that is available to consumers for making decisions. Agnew & Szykman (2005) studies this in the context of information overload, while Phatak (2012) studies the effects of the organization of alternatives on investment decisionmaking.

A set of forthcoming papers also examine the effects of **linguistic** or **textual complexity** on investment decisionmaking. Chin et al. (2021) examines the role of jargon as an impediment to investor understanding of fee concepts. Separately, Scholl et al. (2020*a*) and Scholl et al. (2020*b*) examine the role of textual complexity on mutual fund performance, finding evidence to support the notion of complexity being intentionally curated to obscure weak performance.

Work specifically related to **choice set complexity** - or the difficulty in evaluating tradeoffs between features of items in the choice set - is more sparse in the economics and finance literature, and in particular in the investment decisionmaking literature. This is not surprising in that the preponderance of economic theory has developed with assumptions that embody rational choice, perfect information, and limited search costs. Thaler et al. (2013) briefly discusses ideas related to choice set complexity in the context of trade-offs between disease treatment efficacy and side effects. Agnew & Szykman (2005) use lab experiments to examine choice overload in the context of investment decisions. One of their experimental manipulations involved varying the degree of similarity, with the high similarity condition providing choices in the same Morningstar category with similar strategies and performance metrics; higher similarity increases cognitive burden. The information overload cognitive path in terms of similarity is related to the path proposed for choice set complexity, but the mechanism pursued here is more sharply defined in terms of specific feature tradeoffs that are being made rather than the general similarity of items in the choice set.

As with cardinality complexity, a more developed examination of choice set complexity has been pursued in psychology, marketing and related fields in a number of consumer settings rather than in the economics literature. Tversky & Shafir (1992) provides a fundamental framing of important themes in the context of choice conflict from opposing attribute alternatives, emphasizing the emotional conflicts associated with trading one feature for another. They conduct two studies that help to articulate the issue of cognitive burden associated with tradeoffs between attributes. In the first study, closely related to issue of choice set complexity, they offer betting options or apartment choices. Higher conflict choice sets (those with higher choice set complexity) prompt subjects to procrastinate by requesting more choice options after viewing an initial set. In a second experiment, the authors propose and test a choice environment that ostensibly violates fundamental demand axioms; briefly, if x is preferred to y in the choice set {x,y}, then introduction of choice z that is similar to x and should thus render one option ignorable, can actually create conflict that leads to procrastination.

Other work on the psychological aspects of consumer choice also speak to choice set complexity. Greifeneder et al. (2010) pursues not only the effect of the number of options, but also choice set complexity on consumer decisions. Bettman et al. (1998) provides a wide-ranging review that encapsulates many dimensions of complexity from psychology, decision sciences and marketing in an argument for a constructive theory of consumer choice. In contrast with the familiar textbook economic approach to choice via revealed preferences in which choices reveal underlying pre-existing individual preferences for certain attributes[1], in the constructive view, individuals have a much richer choice process where, for example, there may not in fact be preexisting set of preferences at all. Instead, in some settings, preferences may be developed on the spot as an individual is presented with choice options. The choice environment can stimulate selection of alternative decision heuristics and choice evaluation strategies. Bettman et al. (1998) posits that "[positive] correlation leads to more alternative-based processing for less emotion-laden decisions, because alternative-based processing is likely to elicit negative emotion by highlighting difficult trade-offs." Alternative-based processing considers multiple features of a single choice option before considering other options, while attribute-based processing considers a single attribute for several alternatives before considering another attribute. Such decision processes have received rare consideration in economic applications aside from occasional mentions in multidisciplinary environmental and transportation research (e.g. Hensher (2010)), but these distinctions have maintained at least some appeal in psychology and marketing research on consumer choice (c.f. Dellaert et al. (2019), Mourali & Pons (2009), Jang & Yoon (2016), Fasolo et al. (2009)). Bettman et al. (1993) used simulations and experiments to examine the adaptation strategies used in an effort-accuracy context when responding to choice sets that contain tradeoffs summarized in the correlation structure between features. Widing & Talarzyk (1993) elicited subjects' importance weights and provided a decision aid based on the those weights, which improved accuracy substantially in environments characterized by choice-set complexity. Jia et al. (2004) develop models and provide empirical evidence that suggests higher response error (variations in judgments about the value of a feature or product) is generated when there is more conflict among attributes. Venkatraman et al. (2009) study choice under feature conflict using fMRI data to map the neural paths on decisioning. In more recent work, Haugtvedt et al. (2018)

[1]Bettman et al. (1998) cites to Lucas Jr (1986) as their take on the dominant economic perspective even in the late 1990s, but also admits the interdisciplinary conflict noted by Mcfadden (1997).

explores the role of choice set complexity in other decision framing models, while Fasolo et al. (2009) looks at the implications on the perceived duration of time choosing among alternatives.

The idea and set-up of the second experiment in Tversky & Shafir (1992) is similar to work pursued in the marketing literature on "alignability" of choice set assortment varieties developed by Gourville & Soman (2005). Product assortment varieties (essentially product lines) that are alignable are easily comparable between different alternatives, vis-a-vis nonalignable ones that involve tradeoffs between features. The authors offer examples of alignable product offerings (for example, between different MPG ratings of automobiles), or nonalignable set of product offerings (for example, one car with a sunroof and cloth interior while another car has leather interior and no sunroof, thus forcing the consumer to chose between sunroof and leather interior options). Alignability bears many similarities to choice set complexity, but has typically been concerned with a different unit of analysis and outcomes of interest. Specifically, alignability research tends to be concerned with firm-level outcomes such as market share, and the study of situations where firms introduce additional product lines that lead to a reduction in market share. Yet, the purported mechanism is quite similar: when a firm introduces more nonalignable choices into its product offerings, it may counterintuitively see its market share decline by introducing conflict for within-brand assortments that leads to consumer procrastination, delay or avoidance.

3 Experiment Design

I study the effect of menu set complexity on portfolio allocation behavior using a framework similar to that employed in Choi et al. (2009), as well as recent examples by Beshears et al. (2011), Fisch & Wilkinson-Ryan (2014). To study the effects of complexity, I assigned subjects a menu of five S&P500 index mutual funds and asked each subject to allocate a hypothetical \$10,000 portfolio among the five funds. As the funds are all no-load funds indexed to the S&P 500, the principal differences between the funds are the expense ratio and returns since fund inception. Therefore, the strictly dominant strategy for each subject is to assign the full investment balance to the lowest fee fund on the menu. However, to create a trade-off for subjects, I curated a list of funds that presented some substitutability between fees and past returns for the funds on the menu. This trade-off is the main dimension of complexity that I study.

3.1 Generating Choice Set Complexity

In order to generate random variation while maintaining a comparable menu of options, I restricted the fund selection to a set of twelve S&P 500 candidate funds. Five of these funds were randomly assigned to each subject and were displayed in a random order. These candidate funds have properties and return histories of actual mutual funds, but the details have been anonymized by creating anonymized names (with names based on locations in Chicago), alternative locations for their headquarters (all medium-sized cities in the US), and changing the names of the fund managers. Summary prospectuses were mocked up for the funds using a single template in order to control for any potential variation in the readability of the actual prospectus. Figure 3 provides a full table of the candidate funds, relevant salient information and true names of the fund.

The twelve candidate index funds were selected so that the resulting 792 menu permutations (12 choose 5) would provide a covering of the space on the principal measure of choice set complexity. (More detail on the selection of the candidate funds are provided in the appendix). The main complexity measure I consider in this paper is subject i's assigned menu's intra-menu correlation between expense ratio and its Returns Since Inception (RSI).

$$\rho_i = cor(ExpR_{i,1...5}, RSI_{i,1...5})$$

In the context of this index fund focused choice set, RSI is a somewhat misleading measure of fund performance that is determined almost exclusively by the fund's start date and only creates the appearance of a trade-off for the subject between returns and fees. As such, the set of 12 candidate funds were chosen so that this trade-off would be more or less cognitively difficult for a given subject to assess. The theoretical distribution of intra-menu correlation that is generated from uniform random sampling from the pool of candidate funds is provided in Figure 4. As with our example above, a positive correlation suggests a more complex trade-off between RSI and expense ratio - as RSI increases, the expense ratio also increases; conversely, a negative correlation suggests that for the menu

Fund	Fund name	Return since inception	Fee table
1	HYDE	6.23%	**Annual Operating Expenses (expenses that you pay each year as a % of the value of your investment)** Management Fee 0.20% Other Expense 1.00% Fee Waiver (0.00%) Total Annual Operating Expenses 1.20%
2	ANDERSON	7.09%	**Annual Operating Expenses (expenses that you pay each year as a % of the value of your investment)** Management Fee 0.32% Other Expense 0.35% Fee Waiver (0.03%) Total Annual Operating Expenses 0.64%
3	LAKEVIEW	9.06%	**Annual Operating Expenses (expenses that you pay each year as a % of the value of your investment)** Management Fee 0.25% Other Expense 0.26% Fee Waiver (0.01%) Total Annual Operating Expenses 0.50%
4	ROGERS	7.12%	**Annual Operating Expenses (expenses that you pay each year as a % of the value of your investment)** Management Fee 0.24% Other Expense 0.37% Fee Waiver (0.01%) Total Annual Operating Expenses 0.60%
5	LOGAN	8.60%	**Annual Operating Expenses (expenses that you pay each year as a % of the value of your investment)** Management Fee 0.29% Other Expense 0.44% Fee Waiver (0.28%) Total Annual Operating Expenses 0.45%
6	LINCOLN	8.46%	**Annual Operating Expenses (expenses that you pay each year as a % of the value of your investment)** Management Fee 0.20% Other Expense 0.38% Fee Waiver (0.00%) Total Annual Operating Expenses 0.58%
7	EDGEWATER	7.53%	**Annual Operating Expenses (expenses that you pay each year as a % of the value of your investment)** Management Fee 0.08% Other Expense 0.03% Fee Waiver (0.01%) Total Annual Operating Expenses 0.10%
8	BUCK	9.29%	**Annual Operating Expenses (expenses that you pay each year as a % of the value of your investment)** Management Fee 0.10% Other Expense 0.17% Fee Waiver (0.00%) Total Annual Operating Expenses 0.27%
9	EVANSTON	6.96%	**Annual Operating Expenses (expenses that you pay each year as a % of the value of your investment)** Management Fee 0.05% Other Expense 0.05% Fee Waiver (0.01%) Total Annual Operating Expenses 0.09%
10	WICKER	7.00%	**Annual Operating Expenses (expenses that you pay each year as a % of the value of your investment)** Management Fee 0.10% Other Expense 0.09% Fee Waiver (0.04%) Total Annual Operating Expenses 0.15%
11	PILSEN	7.69%	**Annual Operating Expenses (expenses that you pay each year as a % of the value of your investment)** Management Fee 0.10% Other Expense 0.19% Fee Waiver (0.04%) Total Annual Operating Expenses 0.25%
12	WRIGLEY	5.24%	**Annual Operating Expenses (expenses that you pay each year as a % of the value of your investment)** Management Fee 0.03% Other Expense 0.02% Fee Waiver (0.00%) Total Annual Operating Expenses 0.05%

as a whole, the subject saw higher RSI accompanying lower expense ratios. The histogram of intra-menu correlations that was settled upon was chosen to have a mix of positive and negative values that were somewhat evenly distributed over the range [-1,1]. A number of other menu characteristics were also considered such as average expense ratio, the ratio of the maximum to the minimum expense ratio, the standard deviation of fees and so forth. To consider order issues, the order of funds on the menus was also randomized so that funds appear on different menus in different order.

Figure 5 provides an example screenshot of what a subject would see on their screen for a given menu. Each row in the table represents an anonymized fund name, a link to the fund's mocked-up summary prospectus, returns since fund inception, and finally the fee table.[2] Note that the full statutory prospectus was not provided.

3.2 Characteristics of the Subject Pool

The experiment was deployed on Amerispeak, a nationally representative, probability- based online survey panel developed by NORC at the University of Chicago using an address-based sampling frame. The sampling frame is the same one used for the Survey of Consumer Finances. The Amerispeak panel has an advantage over other probability-level alternatives because it provides better coverage of rural and difficult to access households and greater overall representativeness. The panel also offers clear advantages over Random Digit Dialer approaches that are used in other online survey panels. A sample was generated that allowed completion of the experiment by 4,021 subjects.

Table 1 provides summary statistics for the subject pool on key demographic variables. Column (1) provides the percent of the sample by relevant demographic measure, while column (5) provides weighted averages using survey sampling weights. Unlike student-centered subject pools, the subject pool in this experiment is far more representative of the overall US population with or without weighting-adjustment. As Table 1 indicates, the sample is slightly over-weighted on women (sixty-two percent female), college educated (43

[2]The current paper is part of a larger experimental framework in which several key features are varied on the screenshot provided. These features are easily controlled for by regression adjustment, but are not discussed in detail here in order to simplify the discussion.

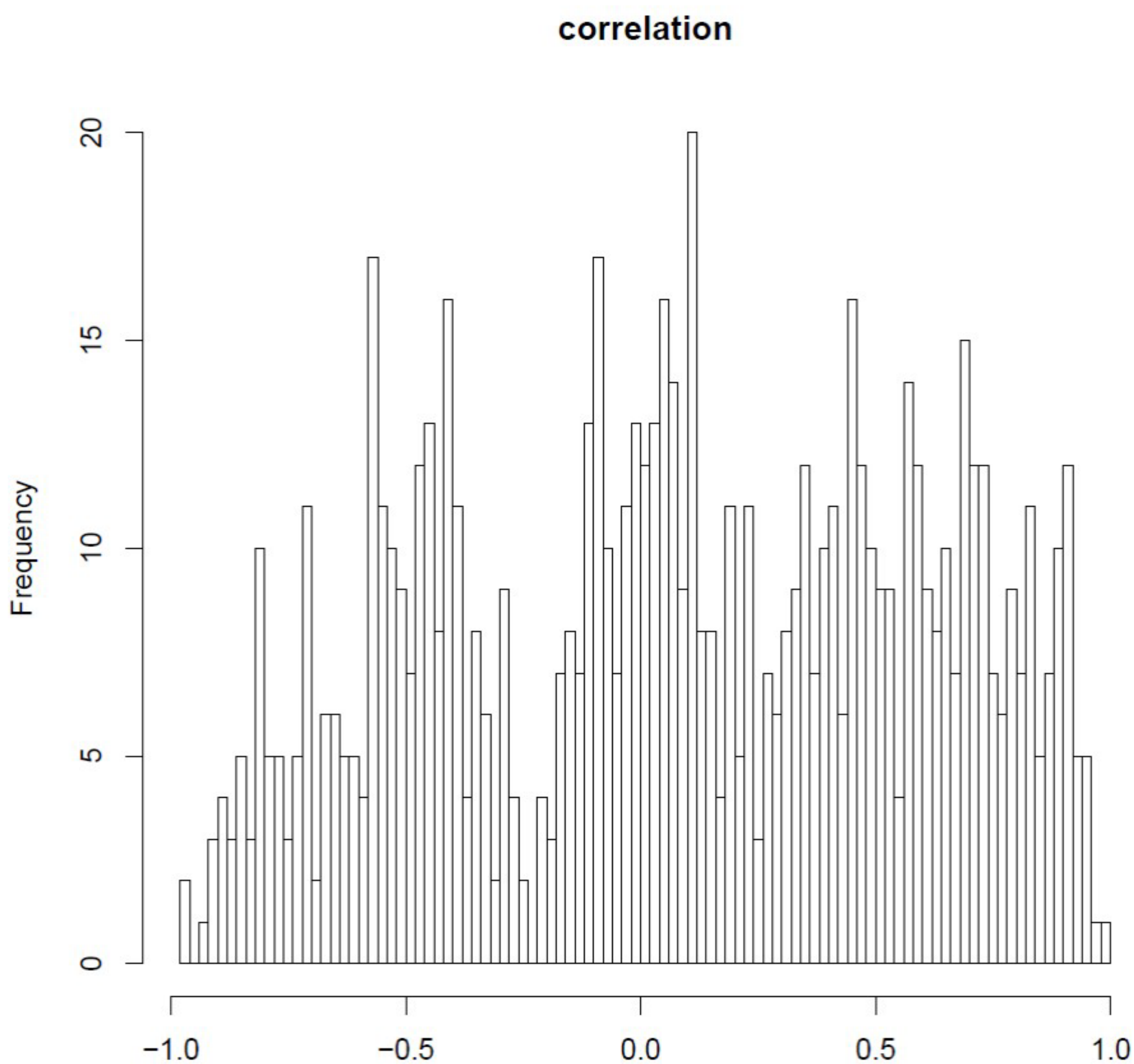


Figure 4: Histogram of Intra-Fund Correlations

Below you should see information on five mutual funds. These are actual mutual funds, but names have been changed.

If you are participating on a mobile device, please hold your device horizontally for easier viewing.

Click the fund name in the Summary Prospectus column to show a prospectus for that fund.

Fund Name	Summary Prospectus	Return since fund's inception	Fee information
ANDERSON	Anderson Summary	7.09%	**Annual Operating Expenses** **(expenses that you pay each year as a % of the value of your investment)** Management Fee 0.32% Other Expense 0.35% Fee Waiver (0.03%) Total Annual Operating Expenses 0.64%
ROGERS	Rogers Summary	7.12%	**Annual Operating Expenses** **(expenses that you pay each year as a % of the value of your investment)** Management Fee 0.24% Other Expense 0.37% Fee Waiver (0.01%) Total Annual Operating Expenses 0.60%
LOGAN	Logan Summary	8.60%	**Annual Operating Expenses** **(expenses that you pay each year as a % of the value of your investment)** Management Fee 0.29% Other Expense 0.44% Fee Waiver (0.28%) Total Annual Operating Expenses 0.45%
LINCOLN	Lincoln Summary	8.46%	**Annual Operating Expenses** **(expenses that you pay each year as a % of the value of your investment)** Management Fee 0.20% Other Expense 0.38% Fee Waiver (0.00%) Total AnnualOperating Expenses 0.58%
EDGEWATER	Edgewater Summary	7.53%	**Annual Operating Expenses** **(expenses that you pay each year as a % of the value of your investment)** Management Fee 0.08% Other Expense 0.03% Fee Waiver (0.01%) Total Annual Operating Expenses 0.10%

Please allocate $10,000 among the five investment options below. You may choose to put your $10,000 to one fund or split your investments among the funds.

- Enter the percentage of $10,000 you would like to allocate to each fund in the table below.
- You may invest in as many OR as few funds as you choose.
- Please be careful to allocate 100% of your funds.

Fund Name	Your allocation (column must total to 100%)
ANDERSON	% 0
ROGERS	% 0
LOGAN	% 0
LINCOLN	% 0
EDGEWATER	% 0
Total	0 %

Figure 5: Menu of Mutual Funds

percent), and the age 30 to 44 cohort. The sample has higher employed levels than the population, but income brackets are largely in line with population estimates from other data sources. Column (8) provides benchmark estimates using the most immediately comparable Survey of Household Economics and Decisionmaking (SHED) or American Consumer Survey (ACS). Comparing columns (5) and (8) suggests that the sampling weights provide a modest, but reasonable representativeness adjustment to the data.[3]

One exciting feature of the fielding of this experiment on a large probability-based survey sample is that it allows conduct an experiment with high internal validity and produce reliable, population-generalizable estimates. Naturally, the results may be specific to the specific experimental set-up. A more complete discussion regarding external validity is below.

[3]Deaton (1997) discusses at length the somewhat religious debate between economic and statistical views on the use of weighting in regression estimates.

Table 1: Summary statistics for experimental subject pool (A1). Means are provided for the experimental sample (unweighted sample mean), the weighted mean of the experimental sample, and estimated population means taken from SHED or ACS external survey data.

variable	mean (sample)	CI	mean (wtd)	CI	mean (pop)	CI	Source
AGE18_29	0.169	(0.157,0.18)	0.171	(0.159,0.183)	0.2029	(19.55,21.03)	SHED
AGE30_44	0.31	(0.295,0.324)	0.264	(0.25,0.277)	0.2506	(24.27,25.87)	SHED
AGE45_59	0.273	(0.259,0.287)	0.264	(0.25,0.278)	0.2567	(24.87,26.48)	SHED
AGE60plus	0.249	(0.235,0.262)	0.301	(0.287,0.315)	0.2897	(28.14,29.81)	SHED
black	0.138	(0.127,0.149)	0.116	(0.106,0.126)	0.127	(0.126,0.12.8)	ACS
collegegrad	0.438	(0.422,0.453)	0.347	(0.332,0.361)	0.30	(0.2991,0.3011)	ACS
employed	0.646	(0.631,0.661)	0.598	(0.583,0.613)	0.6348	(62.59,64.36)	SHED
feesProp	0.307	(0.293,0.321)	0.319	(0.312,0.327)			
female	0.626	(0.611,0.641)	0.51	(0.495,0.526)	0.513	(0.51.2,0.51.4)	ACS
goodinvestor	0.132	(0.122,0.143)	0.123	(0.113,0.133)			
hispanic	0.129	(0.118,0.139)	0.139	(0.129,0.15)	0.183	(0.18.2,0.18.4)	ACS
inc_100_200k	0.18	(0.168,0.192)	0.178	(0.166,0.19)	0.216	(0.215,0.217)	ACS
inc_under100k	0.786	(0.773,0.799)	0.789	(0.776,0.802)	0.709	(0.708,0.71)	ACS
inc200Kplus	0.034	(0.028,0.039)	0.033	(0.028,0.039)	0.076	(-0.075,0.077)	ACS
naive	0.134	(0.123,0.144)	0.139	(0.128,0.149)			
numfunds	3.23	(NaN,NaN)	3.298	(3.247,3.349)			
OWNER	0.419	(0.403,0.434)	0.419	(0.403,0.434)			
retired	0.142	(0.131,0.152)	0.182	(0.17,0.194)	0.1966	(18.92,20.39)	SHED
time_seconds_init_alloc_E2	241.295	(NaN,NaN)	247.654	(NA,NA)			

Table 1 also contains summary statistics on completion of the experimental task. The average value of the main dependent variable feesProp was 0.307. 13.3 percent of subjects were 'good investors,' meaning that they allocated according to the strictly dominating strategy of placing all funds in the lowest cost fund. Table 2 details the percent of the completion subject pool that allocated to the given number of funds. The average subject allocated to 3.23 funds. 40 percent of subjects followed some form of naive diversification, allocating money to each of the five funds offered, but only 13.3 percent of A1 subjects and
15.7 percent of A2 subjects followed a pure naive diversification strategy (equal allocation to all funds).

	V1
1	0.23
2	0.19
3	0.11
4	0.07
5	0.40

Table 2: Fraction of subjects on A1 by total number of funds selected for investment.

3.3 Incentive Compatibility

To create an incentive compatible choice environment and ensure knowledge of the basic terminology of the experiment, prior to viewing the fee menu, all subjects received the instruction screen depicted in Figure 6. This screen notified subjects that better investment performance on the experimental task would be rewarded with a higher payoff. This is in contrast to other choice experiments such as that described in Beshears (2011), which provided each subject with a clearly detailed payoff function as well as an example payoff calculation. However, such studies were conducted in a laboratory environment where such instructions could be reliably provided in person. This was deemed largely infeasible in the current environment, with more potential for confusion than elucidation.

Please read the following information before going to the next page.

- **Mutual Fund**: A mutual fund is a company that takes money from many investors. The fund can invest this money in many ways, including in stocks, bonds, and short-term money-market instruments.
- **S&P 500 Index**: The S&P 500 Index measures the total stock market value of 500 of the largest U.S. companies. It is one of the most widely used stock market indexes in the world. The investment return of the S&P 500 Index is the percent change over time in the total stock market value of these 500 companies.
- **S&P 500 Index Fund**: An S&P 500 Index Fund is a mutual fund that tries to make its investment return (before fees) match the S&P 500 Index's investment return. Funds generally do this matching by investing in stocks in the S&P 500 Index.

Bonus Payment

As part of this section of the survey, you may be chosen to receive a bonus payment. One in every four participants will receive this bonus. The average bonus payment will be approximately 250 points.

The amount of the bonus will be based on the choices you make. This bonus payment is intended to simulate real life investing.

- The amount of your bonus payment is determined by:
  a) The funds you choose
  b) The investment returns of these funds after costs
- The fund options you will see are based on real funds, but the names have been changed. Your bonus payment will be based on the real returns of these funds over the next 90 days.

Figure 6: Instructions

3.4 Randomization Issues

In reviewing results of the experiment, it became apparent that menu assignment randomization did not occur exactly as directed in the research protocol. The original intention was to achieve uniform random assignment of the twelve funds, resulting in 792 choice set combinations. Within each menu, the presentation order of funds was to be assigned randomly, but the design would not attempt to achieve balance on fund order. This approach would have allowed for roughly full support over the [-1,1] range of the correlation distribution, while assigning each particular choice set offering to about five subjects. The protocol would also have allowed for examination of menu order, albeit with incomplete assignment of menu order. That is, not every menu ordering would have been observed within each choice set, but the randomization would be sufficient to reasonably control for order. Instead of this procedure, the assignment process attempted to assign menus for all permutations on funds and fund order. That approach has 12 x 11 x 10 x 9 x 8 = 95,040 permutations; naturally, it is impossible to achieve with only 4,000 subjects. The assignment was made as a lexicographic list rather than a true random assignment as per the original protocol. As a result, only 284 of the intended 792 menus were assigned. In other words, more balance was achieved on menu ordering features than necessary, at the cost of variation in my primary feature of interest.

There are two primary concerns that arise from this incomplete random assignment process. One concern is that the assignment as administered, resulted in incidental correlation (i.e. unintentional correlation) between complexity and subject characteristics. Such incidental correlation could make it more difficult to identify the treatment effect from increased complexity. The second, albeit less problematic, concern is that the distribution that was generated was skewed or does not have enough variance in complexity to uncover a treatment effect, which could make it difficult to make statements about the effects of complexity in general, or could distort estimates. In this section, I describe features of covariate balance and the distribution of complexity. Although incidental correlation and gaps in support are observed to a certain extent for the experiment, their consequence is almost negligible. This is due in part to the random and exogenous assignment of the complexity dosage, and also due to the fact that the support of assignment over complexity space is imperfect, but sufficient to estimate the complexity effect.

Despite the fact that results were unlikely to be affected by the incomplete randomization issues in the current context, to more fully address concerns about the initial randomization, the experiment was re-fielded on a second survey about eight months after the initial trial with an improved randomization procedure following the originally envisioned protocol. I conducted the randomization myself ensuring coverage of the full set of 792 menus, then assigning a random order within each menu. I used multiple methods to randomize assignment. Menus were assigned to subjects at invitation. 2,890 subjects completed the second experiment, including 1,834 of the original subjects and 1056 fresh subjects. All assignments of conditions and menus were made without regard to assignment on the original fielding of the task. This second deployment of the experimental task is herein referred to as A2, while the first administration is referred to as A1. This second administration, employing some of the same subjects as the original experiment, allows for, among other things, fixed effects estimation of the treatment effect.

Complexity Assignment The left panel of Figure 7 provides a pirate plot of complexity assignment for the 4021 subjects that completed the experiment on A1. A pirate plot combines a traditional boxplot with a density plot, plus a bit more information (see Phillips (2017)), with maximum, minimum and median values as per a usual barchart and the width of the blob at any particular y value providing the density of observations in that neighborhood. The bandwidth parameter is set low to highlight a few features of assigned complexity.

A few initial observations are apparent from the pirate plot regarding the distribution of A1. The distribution is slightly skewed towards negative complexity values (i.e. lower levels of complexity), with a median assigned value of corExpRSI is -0.085. The range of corExpRSI values does not extend to the full [-1,1] space, with a maximum value of 0.948 and a minimum at -0.912. It is also apparent that there are several gaps in the support of the distribution, particularly just below -0.2 and around 0.3, but that there is good, if slightly uneven, coverage throughout the distribution. Finally, it is notable that a large fraction of cases were assigned a corExpRSI value in the neighborhood of -0.8.

The right panel of 7 describes a somewhat different assignment pattern on A2. Most

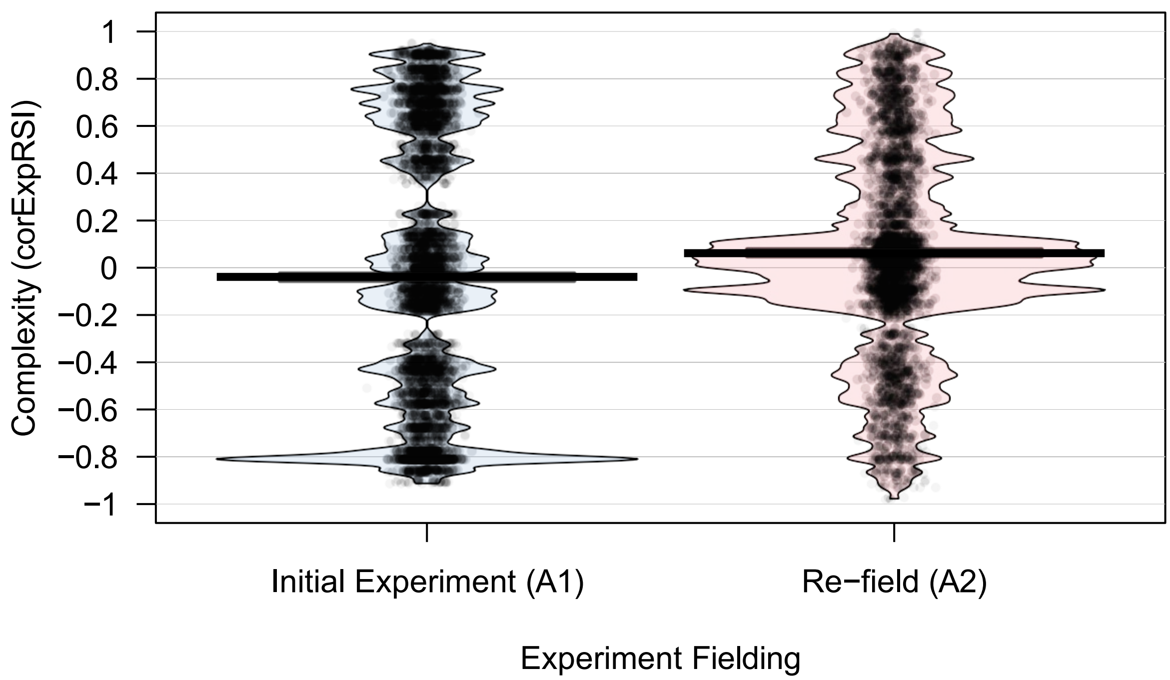


Figure 7: Pirate plot of complexity assignment distribution in initial experiment (A1, left) and re-field experiment (A2, right).

notably, the distribution is skewed more towards higher complexity levels. The median assigned value is 0.037, roughly 12 correlation points higher than in A1. The range is extended as well to [-0.98,0.99]. There is also support throughout the distribution so that much of the complexity range is covered, with much more density and coverage in the positive complexity values. Finally, the bulge in the distribution centers around a complexity value of zero, or more specifically in the range of [-0.2,.15].

The distribution of complexity on A1 suffers from a bit too much negative skewness and small gaps in support, but there do not appear to be overly pressing concerns. The gaps arose idiosyncratically, and neither align with any particular a priori inflection point, nor leave major segments of the distribution unexplored. Even the noted values that were not assigned to any subjects could very plausibly ripe for interpolation. A2 adjusts these anomalies and provides a bit more insight into responses to responses to higher complexity assignments, but the distribution exhibited in A1 appears completely reasonable and is not likely to lead to any compromise of the findings.

Covariate Balance Table 3 provides summary statistics dividing the sample into three discrete groups based on assigned intra-menu correlation values ([-1.0,-0.3];(-0.3,0.3);[.3,1]); these groups were selected to compare three relatively continuous and even segments of complexity space. Visual inspection of Table 3 suggests some differences between the demographic characteristics of the lowest and the highest correlation groups, with the confidence intervals lacking overlap in a number of instances (e.g. female, college graduate, black, and several of the income and age categories). To formally test for balance, Figure **??** provides a boxplot of the p-values from variable-by-variable, group-by-group comparisons for key demographic variables. The fraction of p-values that reject the null of equal means on a given observed variable is not consistent with covariate balance deviations attributable to chance alone, so that assignment of complexity does appear to have created some incidental correlation with demographic covariates. This does appear to present more of a concern than the distributional aspects of complexity assignment, and provides a strong rationale for considering additional covariates in estimation of the complexity effect (see Imbens & Rubin (2015)).

Table 3: Summary statistics for A1 by correlation group. Groups are ([-1.0,-0.3];(-0.3,0.3);[.3,1])

	mean	SE	CI	CI	mean	SE	CI	CI	mean	SE	CI	CI
female	0.612	0.012	0.588	0.637	0.594	0.014	0.566	0.623	0.671	0.013	0.645	0.696
collegegrad	0.457	0.013	0.433	0.482	0.465	0.015	0.437	0.494	0.39	0.013	0.364	0.416
black	0.116	0.008	0.1	0.132	0.148	0.01	0.127	0.168	0.155	0.01	0.136	0.175
hispanic	0.121	0.008	0.105	0.138	0.121	0.01	0.102	0.14	0.144	0.01	0.125	0.163
AGE18 29	0.149	0.009	0.131	0.167	0.17	0.011	0.148	0.191	0.192	0.011	0.17	0.213
AGE30_44	0.291	0.012	0.268	0.314	0.319	0.014	0.292	0.346	0.323	0.013	0.298	0.348
AGE45 59	0.298	0.012	0.275	0.321	0.278	0.013	0.252	0.303	0.239	0.012	0.216	0.262
AGE60plus	0.262	0.011	0.24	0.284	0.234	0.012	0.209	0.258	0.246	0.012	0.223	0.27
employed	0.652	0.012	0.628	0.675	0.664	0.014	0.637	0.692	0.624	0.013	0.597	0.65
retired	0.146	0.009	0.129	0.164	0.147	0.01	0.127	0.167	0.131	0.009	0.113	0.149
inc under100k	0.761	0.011	0.739	0.782	0.782	0.012	0.758	0.806	0.82	0.011	0.799	0.841
inc_100_200k	0.197	0.01	0.178	0.217	0.188	0.011	0.165	0.21	0.154	0.01	0.134	0.173
inc200Kplus	0.042	0.005	0.032	0.052	0.03	0.005	0.02	0.04	0.027	0.004	0.018	0.035
OWNER	0.455	0.013	0.43	0.48	0.452	0.015	0.423	0.48	0.347	0.013	0.321	0.372

Table 4 provides summary statistics by correlation group for A2. With the adjusted assignment of complexity in A2 second fielding, covariate balance for the three complexity groupings examined previously, ([-1,-0.3],(-0.3,0.3),[0.3,1]) are in line with what one would expect from chance alone. That is, confidence intervals for the three groups overlap for the demographic variables (as well as others not presented). Figure **??** provides a boxplot of the p-values for formal two sample t-tests for each variable for each pairing of groups; these suggest only deviations attributable to chance alone.

Table 4: Examination of covariate balance for re-fielded experiment A2. Groupings left to right are [-1,-0.3],(-0.3,0.3),[0.3,1].

	mean	SE	CI	CI	mean	SE	CI	CI	mean	SE	CI	CI
female	0.494	0.008	0.477	0.51	0.527	0.021	0.485	0.568	0.53	0.017	0.496	0.564
collegegrad	0.424	0.008	0.408	0.44	0.401	0.021	0.36	0.441	0.415	0.017	0.381	0.448
black	0.137	0.006	0.126	0.148	0.137	0.014	0.108	0.165	0.147	0.012	0.123	0.171
hispanic	0.146	0.006	0.134	0.157	0.161	0.015	0.131	0.192	0.141	0.012	0.118	0.165
AGE18_29	0.135	0.006	0.124	0.146	0.142	0.015	0.113	0.171	0.128	0.011	0.106	0.151
AGE30_44	0.297	0.008	0.282	0.311	0.285	0.019	0.248	0.323	0.302	0.016	0.271	0.332
AGE45_59	0.284	0.007	0.269	0.298	0.271	0.019	0.235	0.308	0.274	0.015	0.244	0.304
AGE60plus	0.285	0.007	0.27	0.3	0.301	0.019	0.264	0.339	0.296	0.016	0.265	0.326
employed	0.643	0.008	0.627	0.658	0.661	0.02	0.622	0.7	0.641	0.016	0.608	0.673
retired	0.171	0.006	0.159	0.183	0.172	0.016	0.141	0.203	0.17	0.013	0.144	0.195
inc_under100k	0.781	0.007	0.768	0.795	0.764	0.018	0.729	0.799	0.782	0.014	0.754	0.81
inc_100_200k	0.191	0.006	0.178	0.204	0.188	0.016	0.156	0.22	0.183	0.013	0.157	0.209
inc200Kplus	0.028	0.003	0.022	0.033	0.048	0.009	0.03	0.065	0.035	0.006	0.023	0.048

Overall, since choice sets and, in turn, complexity were exogenously and randomly assigned at the individual level, the imperfections in A1 assignment are not a major concern. Nevertheless, regression adjustment as per the procedure described in Lin (2013) provides a basis for correcting these imperfections. Administration A2 offers an opportunity to examine the results with these issues removed entirely.

3.5 Outcomes of Interest

As funds are randomly assigned to menus and not the product of any choice or any personal attribute, it is reasonable to directly observe the relationship of complexity with the outcomes of interest as a preview of the main empirical results. In fact, this approach would be favored by Freedman (2008) and Freedman et al. (2008), because without regression adjustment for additional controls, with randomly assigned experiment conditions a simple comparison of group means provides an unbiased estimate of the Intention to Treat Effect (ITT).

In Figure 9, feesProp increases non-monotonically as choice set complexity increases. The behavior can be best characterized in three domain groupings [-1.0,-0.3), (-0.3,0.25], [0.25,1.0] as identified by the gaps in support at correlations of roughly -0.3 and 0.25. In the lowest (most negative) grouping, the relationship is fairly muted and tends to remain around 0.25. In the middle grouping, feesProp rises fairly directly and almost linearly as the correlation increases. In the final group the level of feesProp has stabilized in the range of .38-.4. A small dip in the non-parametric estimate is observed at the upper correlation range, and a small elevation in feesProp is observed at the very lowest range of complexity, but these nonparametric estimates may be driven by a relatively few observations. In this figure there is suggestive evidence both that higher levels of complexity increase feesProp and that the transition is smooth between high and low correlation ranges. Behavior at the gaps in the complexity distribution is somewhat aberrant.

Figure 9 plots the rank order version of feesProp against the intra-menu correlation. Here a smooth gradual increase is observable, with little evidence of nonlinearities.

In subsequent sections, I provide more formal analysis with conclusions in line with these initial results.

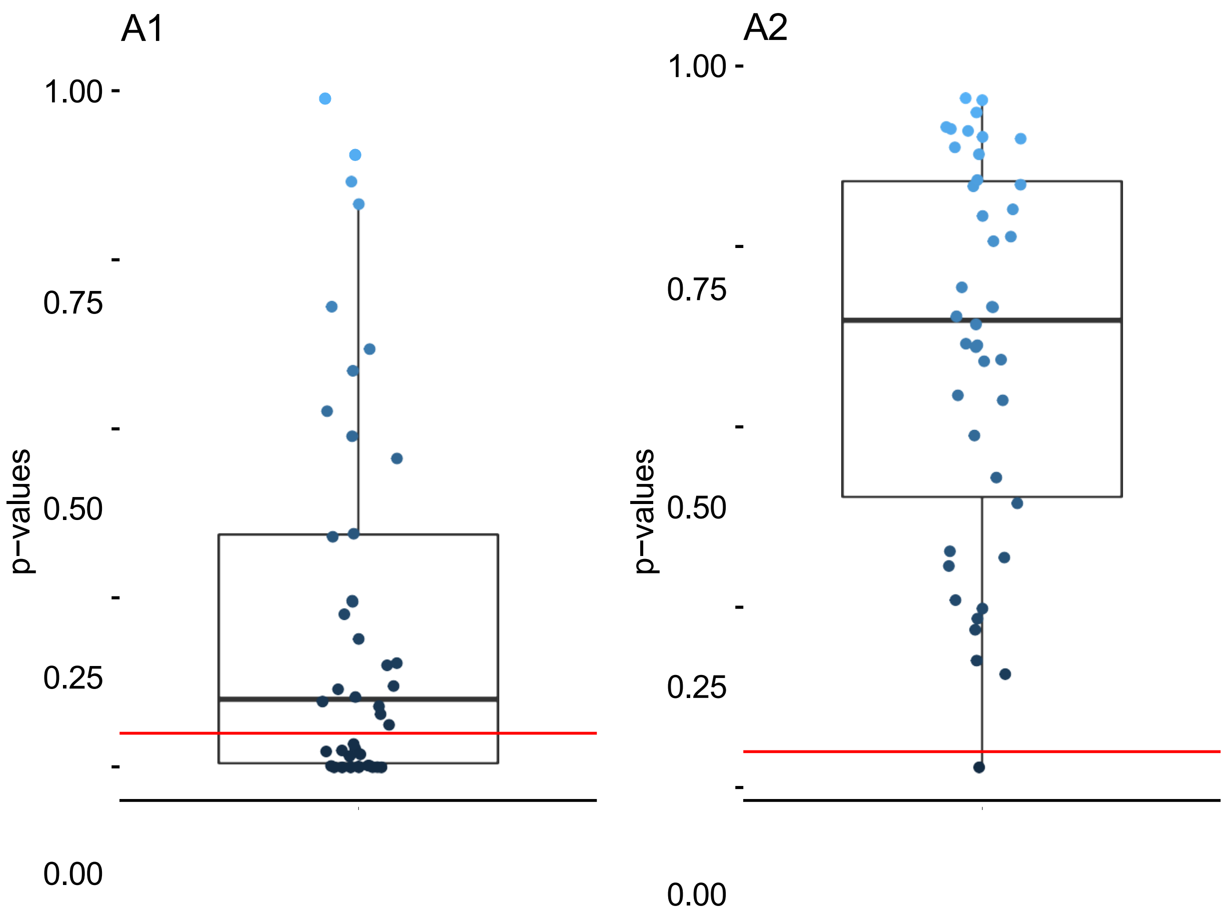
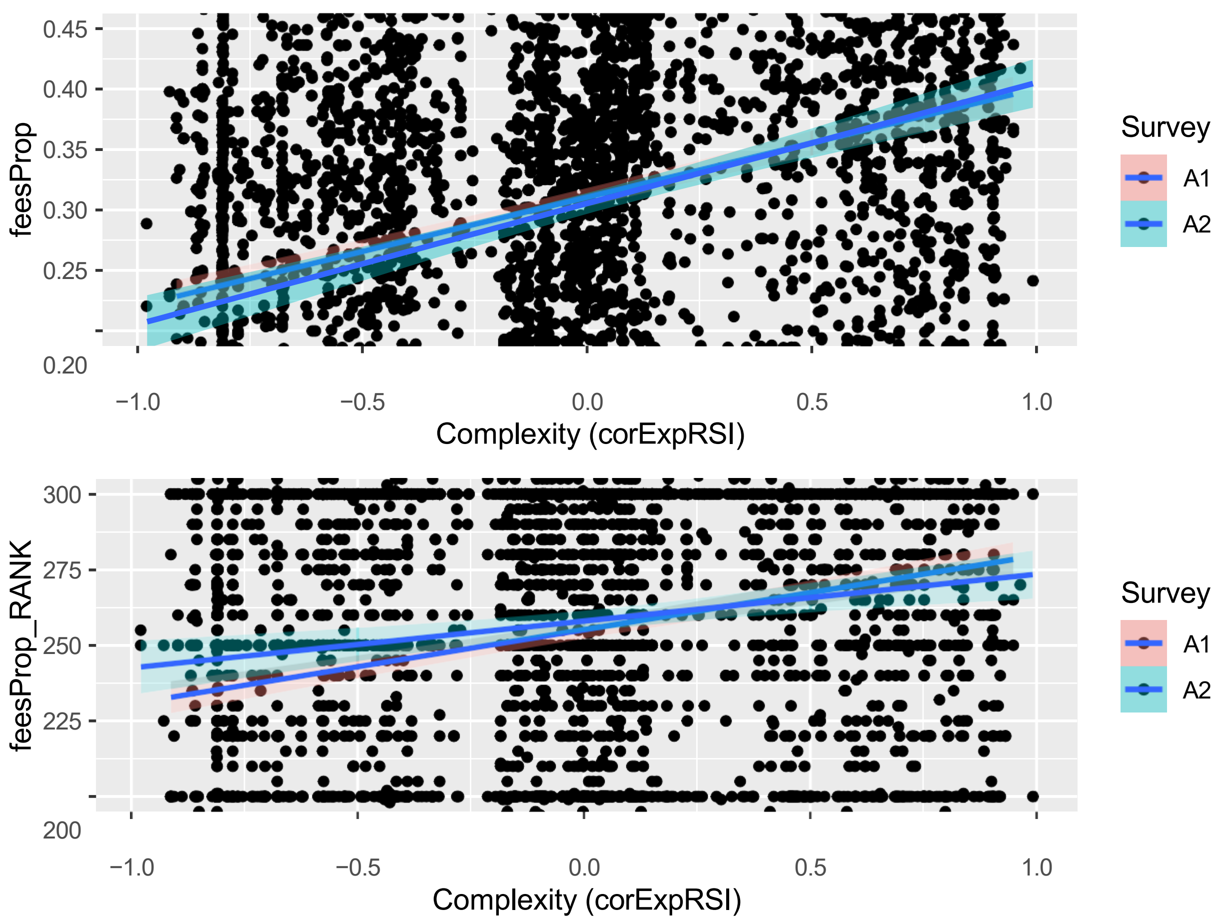


Figure 8: Left panel: Boxplot of p-values of difference in sample means for selected features (A1). Grouping breaks correspond to the intervals [-1,-0.3],(-0.3,0.3),[0.3,1]. Right panel: Boxplot of p-values of difference in sample means for selected demographic features (A2). While randomization in A1 resulted in systematic differences in assignment-group level observables, A2 completely eliminates any covariate imbalances not due to chance alone. Adherence to the randomization protocol in A2 eliminates covariate balance.



0.45
0.40
0.35
0.30
0.25
0.20
feesProp
−1.0
−0.5
0.0
0.5
1.0
Complexity (corExpRSI)
Survey
A1
A2
300
275
250
225
200
feesProp_RANK
−1.0
−0.5
0.0
0.5
1.0
Complexity (corExpRSI)
Survey
A1
A2

Figure 9:

3.6 Econometric Framework

I examine investor choice with summary features of a chosen portfolio. feesProp is the proportion of the maximum total fee the subject could have paid on their assigned menu. That is,

$$feesProp = \frac{FeesPaid - MinFees}{MaxFees - MinFees}$$

Where FeesPaid is the dollar amount of fees for the subject's chosen portfolio, Min Fees (Max Fees) is the dollar amount of fees the subject would have paid if s/he had assigned all funds to the lowest (highest) cost fund on her/his menu. This variable measures the subject's allocation-weighted average of fees on the menu, normalizing on the properties of the menu. Higher values of feesProp indicate higher overall fees paid by the subject.

I pursue a conservative approach to analysis of the results of the experiment in the context a randomization inference setting consistent with the recommendations of Freedman (2008) and Freedman et al. (2008), Athey & Imbens (2017). Such an approach descends from Fisher-Neyman tradition of the analysis of experiments, viewing experimental inference as a missing data problem in the context of potential outcomes, only a subset of which are observable for any particular subject (c.f. Imbens & Rubin (2015)). To say the least, this framework tends to view regression-adjustment to experimental estimates highly skeptically, and, as discussed in Freedman (2008) and Freedman et al. (2008), regression adjustment of randomized experiments can introduce various statistical pathologies; Freedman argues that regression analysis of experiments is not justified in the context of randomized experiments.[4] At the same time, note that many of the concerns raised relate

[4]As Lin (2013) notes, and as reviewed in Imbens & Rubin (2015), alternative perspectives on the analysis of experiments exist. In fact, some treatments such as Angrist & Pischke (2008), discuss the inclusion of covariates almost glibly. Freedman ostensibly was not aware of such work as Tsiatis et al. (2008) and others at the time of his writing. Such perspectives tend to be less conservative about the inclusion of covariates. Imbens & Rubin (2015) articulates experiments in the context of a finite sample experiment with subjects drawn from an infinite superpopulation (hence more of a sampling issue than a missing data problem). In the current context, this perspective is extremely appealing given that the subject pool is a large random sample from a well structured national probability sample. Relaxing the strict randomization inference assumptions is for the most part a matter of interpretation rather than estimation mechanics,

to asymptotic distribution properties of estimators, yet the sample size of nearly 7,000 subjects in the current context is relatively large, so that many of the specific issues raised by Freedman (2008) and Freedman et al. (2008) are not germane.

Despite the tendency to discount regression adjustment in a randomization inference context, there are good reasons for pursuing the practice in this study, in a manner that heeds to the warnings of Athey & Imbens (2017). As described in the previous section, the randomization process was somewhat incomplete, creating some incidental correlation with demographic features as noted in Table 3. Moreover, the subject pool was exposed to other experimental conditions during the course of completing the survey instrument, which could differentially impact conditions for those assigned to different experimental conditions. Finally, there is an interest in the differential implications of complexity for selected demographic and sophistication subgroups.

For these reasons, my results largely follow the careful step-by-step recommendations of Freedman (2008) and Freedman et al. (2008) in order to provide transparency, which includes such tasks as reporting the results of balance tests (as per Table 3 and Table 4). I examine the simple non-regression-adjusted results of complexity on the outcome of interest before cautiously augmenting the estimating equations to introduce additional covariates. Specifically, I develop an approach in the mode of Lin (2013) (also discussed in Athey & Imbens (2017)), which follows the strict randomization inference views of Freedman (2008) and Freedman et al. (2008) - specifically that the N subjects represent the entire population, while not discounting alternative viewpoints to the analysis of experiments.[5] Specifically,

---

as Imbens & Rubin (2015) (see, in particular 7.3-7.9) examines regression adjustment with less strict functional form assumptions than employed here, with concerns mostly related to small sample properties and limits precision improvement from covariates with little predictive value (again, in small samples this can actually have adverse effects on precision, but the samples for A1 and A2 here are each on their own sufficiently large).

[5]As Lin (2013) notes, and as reviewed in Imbens & Rubin (2015), alternative perspectives on the analysis of experiments exist. In fact, some treatments such as Angrist & Pischke (2008), discuss the inclusion of covariates almost glibly. Freedman ostensibly was not aware of such work as Tsiatis et al. (2008) and others at the time of his writing. Such perspectives tend to be less conservative about the inclusion of covariates. Imbens & Rubin (2015) articulates experiments in the context of a finite sample experiment with subjects drawn from an infinite superpopulation (hence more of a sampling issue than a missing data problem). In the current context, this perspective is extremely appealing given that the subject pool is a

Lin (2013) recommends relying on dummy variables that partion the sample, demeaning the covariates, and including all interaction terms. While noting that the precision gains from regression adjustment are typically small, Athey & Imbens (2017) demonstrate that in this specific case the least squares estimator is unbiased for the average treatment effect (see also Imbens & Rubin (2015)).

3.6.1 Identification

To formalize the approach, each subject has a continuum of potential outcomes $Y_i(\rho)$ defined over choice set complexity space $\{\rho : \rho \in [\underline{\rho}, \bar{\rho}]\}$. Within each administration $t \in \{A1, A2\}$ of the one-shot experiment, a random assignment $W_{i,t}$ is made to the subject. Each subject can only be observed at the $\rho(W_{i,t})$ that is a randomly assigned with the subject's choice set, or in other words at a randomly chosen point in the individual's response path (to save notation, I will in places simply refer to this as $\rho_{i,t}$. Here, I pursue an approach that will be relatively familiar in the sense of the usual analysis of causal effects.[6] For simplicity, the interest is in estimating the average causal effect at a given dosage of complexity. [7]

large random sample from a well structured national probability sample. Relaxing the strict randomization inference assumptions is for the most part a matter of interpretation rather than estimation mechanics, as Imbens & Rubin (2015) (see, in particular 7.3-7.9) examines regression adjustment with less strict functional form assumptions than employed here, with concerns mostly related to small sample properties and limits precision improvement from covariates with little predictive value (again, in small samples this can actually have adverse effects on precision, but the samples for A1 and A2 here are each on their own sufficiently large).

[6]The average dose response function would be a slightly different estimate than $\tau(\rho)$, $\mu(\rho) = \mathrm{E}[Y_i(\rho)]$ for $\rho \in [\underline{\rho}, \bar{\rho}]$.

[7]In the current context, set-up may be thought of in a way similar to the usual RCM for binary treatments. That is $\tau(\rho) = \mathrm{E}[Y_i(\rho) - Y_i(0)|\rho = \rho_i]$, where $Y_i(0)$ is the counterfactual as usual, but in this instance it is an estimate of the effect in reference to a counterfactual complexity level of $\rho = 0$ rather than a non-treatment status. Regression estimation is identical in either case. Such formulation is a randomization-inference estimator similar to that of the population average causal effect in Abadie et al. (2014) (equation 2.1), where, in their example, U.S. states are assumed to have a potential outcome in a high and low treatment condition. In principle $\tau(\rho)$ can be defined in reference to any other value of ρ, but stating it formally in the usual Rubin Causal Framework way with a reference value of zero provides a tie to the usual regression framework and the standard estimation of treatment effects. This formulation seemed worthy of mention as continuous treatments remain somewhat unusual in the experimental literature.

I estimate the average treatment effect:

$$\tau(\rho) = \mathrm{E}[Y_i(\rho)|\rho(W_{i,t}) = \rho]$$

Or in a more familiar form similar to the usual potential outcomes framework,

$$\tau(\rho) = \mathrm{E}[Y_i(\rho)|\rho(W_{i,t}) = \rho] - \mathrm{E}[Y_i(\rho)|\rho(W_{i,t}) = 0]$$

This formulation is familiar in the usual sense of estimation of treatment effects as the second term is similar to the usual counterfactual term, except that here the counterfactual is a reference value of $\rho = 0$, rather than assignment to nontreatment status. This yields an estimate of the average treatment effect $\tau = \mathrm{E}[\tau(\rho)]$, which articulates the average change in y for a change in ρ from zero.

The most straightforward approach to estimation of τ by OLS:

$$Y_{i,t} = \alpha + \tau\rho(W_{i,t}) + E_{i,t}$$

,

which is equivalent to the Fisher-Neyman approach of comparing means from different assignment groups. In the setup as formulated, estimate of τ provides an estimate of the effect of a deviation of ρ from zero. For interpretation in the current setting, it is actually more convenient to double this estimate, effectively yielding a comparison of $y(\overline{\rho})$ to $y(\underline{\rho})$, or the change from the lowest level of complexity to the highest. [8]

Introduction of covariates takes the form:

$$\tau(\rho, X) = \mathrm{E}[Y_i(\rho)|\rho(W_{i,t}) = \rho, X = x_i] - \mathrm{E}[Y_i(\rho)|\rho(W_{i,t}) = 0, X = x_i]$$

where X is a set of demeaned indicator variables that partition the population. The treatment effect that is identified is the difference in outcomes for subjects with the same pre-treatment characteristics that have been assigned to a given level of complexity (versus those with the same X values that have been assigned a complexity level of zero). In the

[8]One could alternatively reconfigure identification above to set the reference value of ρ as $\underline{\rho}$ rather than zero, but this adds unnecessary complication. Note that non-parametric estimation could also be pursued $\mathrm{E}[\tau(\rho) - \tau(\underline{\rho})]$.

usual way, the regression estimator employs a (linear) functional form to make interpolation possible for missing values in the covariate space. Using the Lin estimator, this remains an unbiased estimate of the average treatment effect, so that $\tau(\rho) \approx \tau(\rho, X)$, but $\tau(\rho, X)$ has the advantage of a functional specification of X, which enables subgroup analysis as well as providing a basis to project results onto the full population. Then the "augmented" model is then:

$$E[Y_{i,t}| W_{i,t}, X_i] = \alpha + \tau\rho(W_{i,t}) + X^1\beta$$

Which is estimated:

$$Y_{i,t} = \alpha + \tau\rho(W_{i,t}) + \beta^1 X + \eta_{i,t}$$

for $t \in \{A_1, A_2\}$, where the X variables form a partion of the population and have been demeaned. For the "interaction-augmented" approach model advocated by Lin (2013) and Athey & Imbens (2017) the estimator becomes:

$$Y_{i,t} = \alpha + \tau\rho(W_{i,t}) + \beta^1 X_{i,t} + \gamma^1(\rho(W_{i,t})X_{i,t})) + \eta_{i,t} \tag{1}$$

which provides an unbiased estimate of the average treatment effect. In Equation 1, β provides the level change in the outcome, which is often interesting, but not particularly germane in the current paper. Rather, γ provides an easily interpretable solution as the difference in the average treatment effect by subgroup.

Despite the notation, the analysis thus far does not exploit the longitudinal structure. That is, analysis of A1 and A2 each provide separate estimates of $\tau(\rho)$ and $\tau(\rho, X)$. Estimation by pooled panel regression is straightforward by applying 1 either in a balanced panel (only using subjects for which $i \in \{A_1\}\cap\{A_2\}$) or unbalanced panel ($i \in \{A_1\}\cup\{A_2\}$). The administration of A2 on 1,834 of the 4,021 original A1 subjects, provides the opportunity for a within-subjects estimate:

$$\tau^{FE}(\rho, X) = \mathsf{E}[Y_{i,A2}(\rho(W_{i,A2})) - Y_{i,A1}(\rho(W_{i,A1}))|i \in A_1 \cap A_2]$$

Fixed effects is particularly advantageous because within subjects identification reduces some concerns about the incidental correlation with demographics in A_1 as fixed demographic terms drop out. The fixed effects estimate of $\tau(\rho, X)$ is provided by:

$$Y_{i,t} = \alpha_i + \tau\rho(W_{i,t}) + \beta^1 X_{i,t} + \gamma^1(\rho(W_{i,t})X_{i,t})) + \varepsilon + \eta_{i,t} \tag{2}$$

In the base and augmented versions of the fixed effects model, β and γ are set to zero, respectively.[9] As usual, pre-treatment variables drop out of 2, but other variables such as design effects do not. Because $W_{i,A1} \perp W_{i,A1}$, the induced variation in $\rho(W_{i,t})$ for a subject in both A1 and A2 permits interaction with pre-treatment characteristics. Naturally, with T=2 fixed effects is identical to the first differences estimator, which has an intuitive interpretation here as the estimate of τ expected change in Y given an associated change in ρ.

While a within-subjects approach may create learning from experience in some settings, it is unlikely to be the case in the current context. First, subjects were not provided with direct feedback on their performance in the experiment, so there is unlikely to be learning. Second, anecdotally, the survey administrators suggest that recall on this particular survey platform does not appear to be very strong given the variety of topical exposure that respondents receive. Finally, subjects were directly asked if they recalled similar questions (of which A2 constituted a single question) being asked before - only a third of the subjects that took both A1 and A2 showed any familiarity with any part of the survey instrument on which A2 was administered. (By comparison note that 21 percent of fresh subjects drawn for A2 stated that they had been asked some or almost all of these questions previously).

4 Results

[9]In general, the fixed effects estimate of $\tau\ (\rho, X)$ will be comparable to pooling and OLS estimates in the context it is used here. I tested a variety of OLS, pooling and fixed-effects base specifications (unaugmented) in conditions similar to the set-up of the experiments of this study. On average, the differences in ability to recover the true treatment effect were trivial.

		(1) A1 only	(2) A2 only	(3) A1+A2, Fixed Effects
			Dependent Variable: feesProp	
Panel A	$\tau(\rho)$	0.090***	0.100***	0.097***
NULL Model		(0.007)	(0.010)	(0.009)
	N	4.021	2890	6911
	R^2	0.044	0.034	0.058
	$\tau(\rho)$	0.078***	0.103***	0.098***
Panel B		(0.006)	(0.010)	(0.009)
Augmented	N	4.021	2890	6911
	R^2	0.131	0.114	0.075
	$\tau(\rho)$	0.081***	0.093***	0.098***
Panel C		(0.007)	(0.012)	(0.010)
Interactions	N	4.021	2890	6911
	R^2	0.140	0.128	0.105
	Note:	*p<0.1	**p<0.05	***p<0.01

Figure 10: Consolidated estimation results for outcome feesProp. Panel A - No regression adjustment; Panel B - includes demeaned covariates; Panel C includes demeaned covariates and all interactions with complexity. Columns provide estimates for administration A1 (column 1), administration A2 (column 2), and fixed effects estimates using A1 and A2 data (column 3)

Table 10 presents estimates of the complexity effect on feesProp with three different estimation specifications (row panels) for and three data samples (columns). Panel A (first row) provides estimates from the base specification with only ρ as a regressor. Panel B (second row) provides estimates from the augmented model, which adds to the base model a selection of demeaned covariates. Panel C (third row) adds to the augmented model all interactions of the demeaned covariates with complexity. Column (1) provides estimates using data from A1 only, column (2) provides estimates using A2 only. Column (3) is slightly different approach in that it presents a fixed effects version of the specifications used in Panels A, B, and C, using only the 1,834 subjects that participated in both A1 and A2; while the estimator and sample are different, interpretation of the coefficient values in terms of the complexity effect on feesProp remains comparable to the other estimates. Each panel-sample combination presents the point estimate on complexity and its standard error, the number of observations, and the estimate's adjusted R^2.

Each estimate is constructed in consistency with the conservative randomization inference perspective on the analysis of experiments. The estimates are of the average causal effect $\tau(\rho)$, in line with the comparison of group means developed in the Fisher-Neyman approach to the analysis of experiments. Panel A's results, with no regression controls, are the approach favored in Freedman (2008) and Freedman (2008). As previously discussed, operationalization of the random assignment mechanism in A1 resulted in incidental correlation of ρ with observable and unobservable characteristics. Although menus were randomly assigned at the individual level, this incidental correlation creates both a lack of covariate balance and a potential confound between complexity and subject characteristics. Panels B and C present alternative remedies to address the potential confoundedness issue using the augmented models discussed in Athey & Imbens (2017) and Imbens & Rubin (2015). The approach in Panel B allows for level shifts in feeProp, while Panel C further augments the estimates of Panel B with interactions of the complexity measure with all demeaned covariates. Estimates in Panel C provide a version of the "agnostic" randomization-inference consistent estimation approach advocated by Lin (2013). Unlike the augmented model of Panel B, Lin's approach provides an unbiased estimates of the average treatment effect (Athey & Imbens (2017)).

The estimate of $\tau(\rho)$ on A1 is 0.09. Strictly speaking, the estimates of $\tau(\rho)$ provided are the estimates of $\tau\,(\rho = 1)$, but this is simply the usual interpretation of $(\tau\,(\rho = 1) - \tau\,(\rho = 1))/([\rho = 1] - [\rho = 0])$ in reference to an increase from a reference complexity level of zero to a complexity level of one. Thus, as the relevant concept is going from a low complexity state ($\rho = -1$) to a high complexity state ($\rho = +1$), relevant interpretation requires doubling of the coefficient so that a move from the lowest levels of complexity to the highest levels of complexity in the menu offerings that I provide, results in a 18 percentage point increase in feesProp. The model's R^2 is 0.04, suggesting that complexity alone explains a modest level of the overall variation in fee avoidance behavior. A2's better implementation of the randomization protocol that results in less incidental correlation with observable covariates attributes a slightly higher effect size of 0.10, while the R^2 does moderate slightly. The fixed effects estimate is 0.97, slightly closer to the A2 result; this estimate has a slightly different interpretation as the within-subjects effect, indicating that subjects that received a one-unit higher (lower) level of complexity in the second experiment on average constructed a portfolio with a 0.097 increase (decrease) in the level of feesProp. Put plainly, subjects that were assigned higher levels of complexity in the second experiment, on average fared worse, while subjects that were assigned a lower level in the second experiment performed better. Note in the fixed effects specifications, only 3,668 observations are actually used - two observations for each subject that participated in both A1 and A2; of course, with T=2, the fixed effects specification is the first difference estimator. The relationship of the within-subjects effect can be observed most clearly in Figure 11, which plots an OLS fit of the first difference estimator (without additional controls) for the subjects that participated in both studies.[10]

Panel B of Table 10 augments the simple regression with demeaned covariates. The covariates introduced in the specifications in this table include design elements (features of potentially confounding random assignment experiments), and demographic variables (gender, education groups, retired, African American, Hispanic, age groups). Via Athey & Imbens (2017) the estimates from this model are not necessarily unbiased. In column (1) the estimate on A1 is indeed about 12 percentage points lower for $\tau(\rho)$ than in the baseline

[10]Pooled estimates, not shown, using either the unbalanced sample (all subjects) or unbalanced sample (only repeat subjects), produce quite similar estimates to those presented in Panels A, B and C.

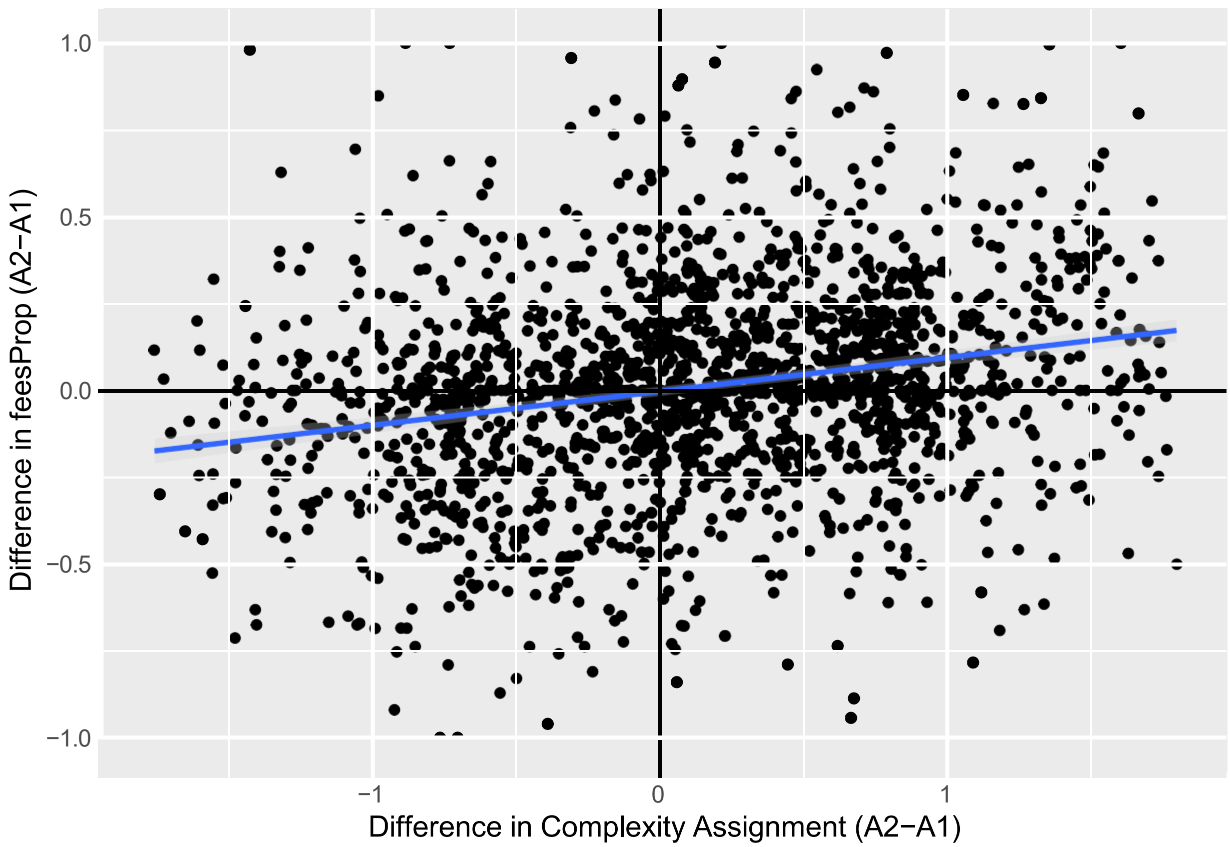


Figure 11: First difference illustration of the complexity effect for subjects completing both A1 and A2.

model, but this is only a moderate difference from the results in Panel A. Column (2) results for A2 show an extremely slight inflation of the point estimate from 0.100 to 0.103. Column (3) represents no change from the Panel A estimate. Note that the fixed effects specifications here do not in fact drop all covariates. The demographic variables are largely absorbed by the fixed effect, but the design elements are not because any relevant design features were reassigned in the second administration of the experiment. The R^2 improves slightly in the fixed effects specifications in Panel B, but there is a marked increase in R^2 in the estimates for A1 and A2. For several of the demeaned covariates, test statistics reject the null of no effect in some instances, representing a level shift in fees paid by group, but not a change in the complexity effect.

Panel C provides estimates for the complexity effect with the specification that includes all the demeaned covariates of the augmented model, along with interactions of the mean-adjusted covariates with complexity. This estimate was originally proposed by Lin (2013) using a randomization inference approach in response to Freedman (2008) and Freedman (2008), is unbiased by construction (see also Athey & Imbens (2017)), and rests firmly within the assumptions of Freedman's critique of the introduction of covariates into the analysis of experiments. When covariates are introduced into the regression, the model's fit improves considerably, while there is very little change in the point estimates on the parameter of interest. Estimates for A1 demonstrate some extremely slight moderation of the complexity effect for A1 from the baseline specification, with to 0.079 from 0.090. The estimate for A2 is at 0.102 from the original estimate of 0.100. The R^2 for these specifications are indeed considerably higher as compared with the R^2 of Panel A. Column (3) of Panel C demonstrates a similar pattern: the point estimate on complexity changes little, but demeaned design elements and their interactions with complexity improve the overall explanatory fit of the specification. Interaction effects in the models in Panel C are discussed below, but are largely unsupportive a differential effect by group.

Overall, results across the specifications in Panels A, B and C, and with samples A1, A2 and the combined sample exhibit remarkably consistent results. The Freedman critique is ultimately of little consequence in the current setup. Estimates provided by the specifications in Panel B are nearly identical to estimates obtained when the same regression controls and interactions are included without Lin's adjustment (not shown). At the same time, this observation only holds in the current case for the model with no interactions. Estimation of interaction effect of Panel C without proper covariate adjustment yields substantively different estimates for both A1 and A2. In the present case, if anything, the bias from introduction of the covariates attenuates the estimate of the treatment effect, but it only results in a consequential change in an estimate of τ (ρ) in specifications in which interaction effects are included.

Anova results in Table 5 (corresponding to the model presented in Panel C, column (1) of Table 10 provides additional insight. While covariates do explain about ten percent of the overall variance, they do so at the expense of 21 degrees of freedom. In contrast, complexity alone describes a sizable fraction of the variance in both A1 and A2 (not shown).

Table 5: ANOVA table for A1, model includes demeaned covariates and interactions.

	Df	Sum Sq	Mean Sq	F-value	p-value
Complexity	1	11.0327	11.0327	202.8261	0.0000
Covariates	21	21.9821	1.0468	19.2439	0.0000
Interactions	21	2.3364	0.1113	2.0454	0.0033
Residuals	3977	216.3280	0.0544	NA	NA

4.1 Robustness

Results presented in Table 10 are robust to alternative specifications. Table 12 provides robustness results for A2 with the following augmented regression results: column (1) is the result previously reported in Panel C (i.e. regression using A2 data with all covariates and interaction terms), Column (2) adds to this specification an indicator for mutual fund ownership and the mutual fund knowledge score of Scholl & Fontes (2022); column (3) adds to the specification of column (2) the subject's investment amount in the best fund on A1; column (4) replaces corExpRSI with an indicator variable for corExpRSI>0. The specification in column (2) results in loss of about a third of the sample because the

	Dependent variable:			
	feesProp			
	(1)	(2)	(3)	(4)
corExpRSI	0.102***	0.110***	0.109***	
	(0.010)	(0.012)	(0.012)	
corExpRSI_Positive				0.074***
				(0.008)
F Statistic	9.729***	9.258***	10.49***	8.407***
Observations	2,890	1,834	1,834	2,890
R^2	0.128	0.189	0.216	0.113
Note:		*p<0.1; **p<0.05; ***p<0.01		

Figure 12: Robustness checks using alternative specifications using experimental sample A2. Column (1) is the interactions-augmented specification presented in Figure 10 column (2), Panel C; column (2) augments the model of column (1) with covariates for mutual fund ownership and the Scholl-Fontes mutual fund knowledge score; column (3) adds the subject's investment amount in the best fund on the A1 experiment for subjects that were in both A1 and A2; column (4) provides the most direct application of the Lin estimator with a binary treatment indicator for Complexity>0.

ownership and knowledge data was collected only on the A1 instrument, and in the column (3) estimation only uses subjects that took the experiment in A1. Despite the loss of sample, the coefficient estimates are largely unchanged.

Column (4)'s step-function approach is suggestive of a slightly attenuated complexity effect, but this is a different estimate entirely as the difference in average feesProp between negative and positive complexity regions. This is technically the estimator that arises without modification from the Lin framework using a binary treatment variable. When both this indicator variable and the continuous corExpRSI are included, the indicator is insignificant and the point estimate on corExpRSI is similar to estimates provided earlier.

4.2 Subgroup Complexity Effects

The interactions effects that were suppressed for A1 in earlier presentations of the model in Panel C Column (1) of Table 10 is presented in Column (1) of Table 13; subsequent columns augment this base model with additional interaction terms, including mutual fund ownership status, Scholl-Fontes mutual fund knowledge scores, and other values (results for administration A2 are largely similar). Column (2) introduces a mutual fund owner dummy and the Scholl-Fontes mutual fund knowledge score; column (3) drops education; columns (4) and (5) introduce discrete dummies for Scholl-Fontes score ranges. Because the task here is not to identify the main complexity effect, but subgroup interactions, point estimates and standard errors are as presented previously, but the asterisks signifying significance have been adjusted with a Benjamini-Hochberg false discovery rate (FDR) adjustment to avoid malfeasance.

A few observations are immediately apparent from the results presented in 13. First, the coefficient on the complexity effect is largely unchanged across specifications from the value reported previously. Second, there is very little evidence that any of the sub-groups implied by these interaction terms demonstrates any modulation of the complexity effect. Without the FDR adjustment, a few of the education and age groups were significant at the 95 percent confidence level, but these were fairly inconsistent across specifications and the point estimates vary widely; with the FDR adjustment, the only coefficient that demonstrates a week effect is the 18-24 age group dummy. Third, including these other interaction terms leads to some slight improvement in the fit of the regression, raising the R^2 by as much as 37 percent. Note that these additional covariates were not included in the base specification because of the desire to keep a consistent specification across A1 and A2 samples - several of these covariates are not available for many A2 observations leading to unfavorable sample loss. Finally, the one covariate for which one can reject the null of no effect is a dummy indicator for having an Scholl-Fontes score above 6. According to the specification in column (5), this interaction has little impact on the complexity effect. For this subgroup, the estimates suggest a complexity effect of approximately 0.018 - and the FDR-adjusted p-values for this term reject the null of no effect at the 95 percent confidence level. According to these results, a very high level of mutual fund knowledge may reduce

- although not completely eliminate - the complexity effect. At the same time, only 6.5 percent of participants in A1 fared this well on the Scholl and Fontes score, echoing the concerns of Scholl & Fontes (2022) that high levels of knowledge are quite rare.

	Dependent variable:				
	feesProp				
	(1)	(2)	(3)	(4)	(5)
complexity	0.079*** (0.006)	0.073*** (0.006)	0.074*** (0.006)	0.073*** (0.006)	0.076*** (0.006)
female	−0.005 (0.013)	−0.007 (0.013)	−0.007 (0.013)	−0.007 (0.013)	−0.011 (0.013)
no HS diploma	0.008 (0.033)	0.015 (0.034)			0.010 (0.033)
HS graduate	−0.005 (0.020)	−0.004 (0.020)			−0.007 (0.020)
some college	0.005 (0.015)	0.005 (0.015)			0.002 (0.015)
age: 18-24	−0.081 (0.047)	−0.106* (0.046)	−0.102* (0.047)	−0.108* (0.046)	−0.101* (0.046)
age: 25-34	−0.061 (0.042)	−0.078 (0.042)	−0.072 (0.042)	−0.075 (0.042)	−0.080 (0.042)
age: 35-44	−0.035 (0.042)	−0.056 (0.042)	−0.052 (0.043)	−0.055 (0.042)	−0.054 (0.042)
age: 45-54	−0.040 (0.043)	−0.058 (0.042)	−0.052 (0.043)	−0.057 (0.042)	−0.059 (0.042)
age: 55-64	−0.052 (0.040)	−0.066 (0.040)	−0.064 (0.041)	−0.068 (0.040)	−0.069 (0.040)
age: 65-74	−0.021 (0.038)	−0.031 (0.038)	−0.027 (0.039)	−0.030 (0.038)	−0.029 (0.038)
black	−0.002 (0.019)	−0.008 (0.019)	−0.008 (0.019)	−0.011 (0.019)	−0.003 (0.019)
hispanic	−0.013 (0.020)	−0.008 (0.019)	−0.005 (0.019)	−0.002 (0.019)	−0.013 (0.019)
retired	−0.040 (0.026)	−0.046 (0.026)	−0.045 (0.026)	−0.046 (0.026)	−0.050 (0.026)
m.f. owner		0.0003 (0.015)	0.002 (0.015)	−0.00002 (0.015)	0.0002 (0.014)
SF score		−0.004 (0.003)	−0.004 (0.003)	−0.004 (0.003)	
sub. numeracy: 3-6				−0.005 (0.022)	
sub. numeracy: 7-11				0.012 (0.014)	
obj. numeracy				−0.0004 (0.007)	
SF score > 6					−0.058** (0.020)
F Statistic	15.114***	19.646***	21.294***	20.185***	19.51***
Observations	4,021	4,021	4,021	4,011	4,021
R^2	0.140	0.189	0.180	0.193	0.188
Note:				* p<0.1; ** p<0.05; *** p<0.01;	

Figure 13: Mean-adjusted interaction effects for sample A1. Column (1) presents results of the main interaction effects of the primary specification from the model presented in Figure 10 Column (1), Panel C, using the complete Lin-style estimator. Subsequent columns introduce additional interaction terms. Significance levels adjusted by Benjamini-Hochberg false discovery rate adjustment.

4.3 Placebo Test

As a placebo test, thirty-six percent of subjects in each of A1 and A2 were randomly assigned to a menu screen that did not provide returns since fund inception on the allocation screen. Therefore, for these subjects, the only salient information that the subjects received was regarding fees of the funds. Returns since inception continued to be provided on the summary prospectus, however subjects rarely opened up the prospectus. Only 22 percent of respondents in A1 clicked on one or more prospectuses. As such, while this did not change the underlying complexity of the choice items, the placebo treatment did not highlight the key features that provided a basis for the tradeoff to the subjects.

Figure 14 illustrates the effect on feesProp for subjects that received the RSI presentation vs those that were assigned to the placebo condition. Subjects that did not receive salient RSI information were only modestly affected by menu set complexity. When features that embody the complexity of the choice set are not present, the trade-off between these characteristics is largely not apparent, and it does not affect the choice being made by subjects. Formal estimates are provided in Table 15. Inclusion of salient RSI does not affect the main complexity effect (indeed, these are the previously reported models from Table 10). In fact, the model suggests that the complexity effect on feesProp may be higher than previously reported for the subgroup that received salient RSI information.

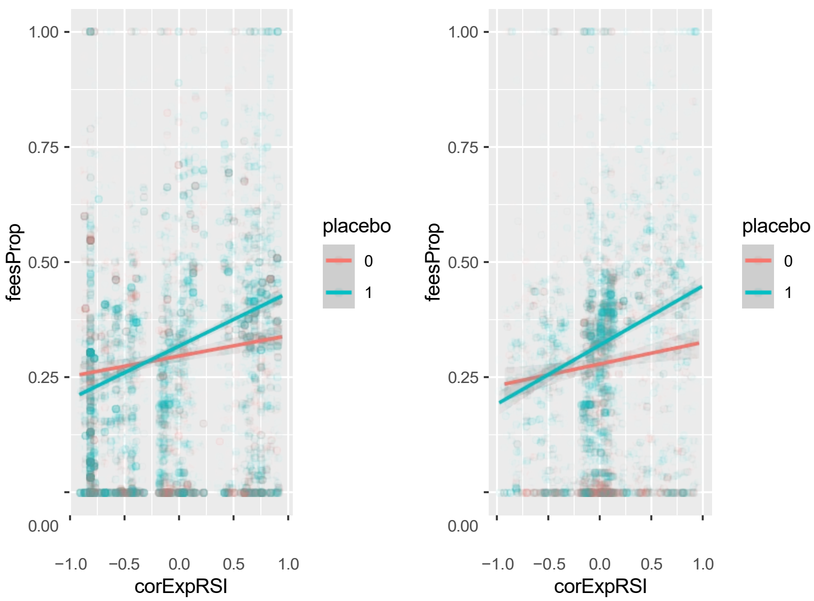


Figure 14: Placebo test of dropping RSI salience.

	Dependent variable:	
	feesProp	
	(1)	(2)
corExpRSI	0.079*** (0.006)	0.102*** (0.010)
exp2rsi.demean	0.012 (0.012)	0.040*** (0.012)
corExpRSI:exp2rsi.demean	0.073*** (0.020)	0.067** (0.028)
F Statistic	15.114***	9.729***
Observations	4,021	2,890
R^2	0.140	0.128
Note:	*p<0.1; **p<0.05; ***p<0.01	

Figure 15: Placebo-test results from interaction of complexity with mean-adjusted indicator for RSI salience and its interaction with complexity for samples A1 and A2. This table presents results from previously suppressed regressors in Panel C of Figure 10

5 Decision Elements and Mechanisms

Results thus far indicate a robust complexity effect that has adverse consequences for the consumer, with very little evidence that complexity moderates across subgroups. In this section, I explore several aspects of subjects' choices that provide insight into the decisions subjects made. In the first subsection, I explore how subjects' choices changed within their menu as complexity increased. In the second subsection, I attempt to classify the allocation strategies that subjects pursued. In the third subsection, I explore how these allocation choices map to potential mechanisms. Finally, I discuss potential issues of recall and learning for repeater subjects that participated in both A1 and A2.

5.1 Allocation Effects of Complexity

How did subjects' choices change in response to higher complexity? The outcome variable feesProp helps to summarize the overall allocation and contextualize it in terms of welfare, but it is not particularly helpful to understand how complexity affected choices. To gain richer insight on these choices, I focus on a different set of outcomes that add more context to actual allocation choices. In particular, I define the variables: fRank1,. . . fRank5 as subjects' allocation to funds ranked from cheapest to most expensive; N Funds is the number of funds to which the subject allocated; and Herfindahl is the usual index of concentration applied to the portfolio allocations.

Results for allocation within the menu according to a fund's fee rank are provided in Figure 16. The coefficients indicate that as complexity increases, subjects allocated less to the cheapest two funds on the menu (Fee Rank 1), and shifted that allocation to the more expensive funds on the menu (by definition, the coefficients sum to zero). Table **??** provides regression results for the outcomes Herfindahl and N Funds. One might imagine that one route through which complexity adversely affects allocation efficiency is by prompting subjects to naively diversify when the choice set is difficult to evaluate; these results do not suggest that is the case. These results reinforce the fact that complexity worsens fee avoidance, and it does so by prompting subjects to an even gradation of flows from low fee funds to high fee funds.

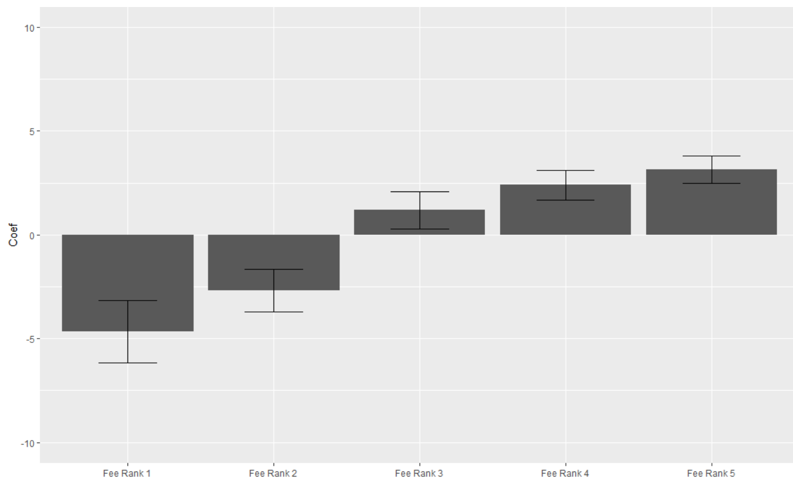


Figure 16: Complexity effect coefficient estimates based on individual funds' rank on the investment menu. Fee Rank1 is the lowest cost fund on the menu, Fee Rank5 is the highest cost fund on the menu. Regressions do not include other covariates (results do not substantively change with the inclusion of covariates).

Table 6:

	Dependent variable:	
	Herfindahl	N Funds
	(1)	(2)
Complexity	-0.028^{***}	0.170^{***}
	(0.009)	(0.044)
Observations	4,021	4,021
Residual Std. Error (df = 4011)	0.324	2.438
Note:	*p<0.1; **p<0.05; ***p<0.01	

C	fRank1	fRank2	fRank3	fRank4	fRank5	rRank1	rRank2	rRank3	rRank4	rRank5	N funds	N	Strategy
1	41.95	41.38	9.04	6.00	1.63	5.12	3.26	71.20	4.94	15.47	2.05	518	Diversified Fee Avoi
2	5.43	6.01	6.34	9.30	72.93	15.86	12.31	8.69	33.96	29.18	2.45	386	Fee Confused
3	62.27	19.90	9.30	5.54	3.00	33.32	35.89	4.00	3.00	23.78	2.49	626	Diversified Fee Avoi
4	18.14	21.80	21.79	21.50	16.78	20.88	24.44	19.93	18.08	16.66	4.44	3350	Pure Naive Diversifi
5	94.61	3.32	1.28	0.48	0.31	1.41	0.98	1.76	1.24	94.61	1.35	643	Optimal Strategy
6	62.22	27.43	4.71	4.97	0.66	6.53	3.33	2.33	77.30	10.50	1.80	479	Diversified Strong F
7	3.50	30.07	42.03	22.55	1.84	72.09	15.54	7.49	2.79	2.09	2.21	909	Returns Chasing

Table 7: Cluster analysis of subjects' allocation strategies.

5.2 Allocation Strategies

For economists, choice contexts are typically somewhat rigid. The foundation of the revealed preference approach to choice requires consumers to have a pre-existing set of preferences that may be elicited by observing their choices in certain settings. Thus, if a consumer avoids a risky financial product, they might provide evidence of lower risk tolerance. Other assumptions may be embedded in the typical model such as some form of rationality, attention, specific utility formulations and so forth, but the standard view is largely that these are modeling assumptions that capture inherent characteristics of the individual.

While this perspective is omnipresent in economics, it is actually quite different from some views that exist in other fields. In psychology and marketing, some literature has suggested the possibility of *constructed preferences* (Bettman et al. (1998), Haugtvedt et al. (2018), Hsee & Zhang (2010)); from this view,a consumer may not be solving an optimization problem based on pre-exiting preferences, but rather she formulates her preferences on the spot when presented with a particular choice context. For example, a consumer may not go to buy a TV with a pre-existing sense of pixels, dimensions or other features, but may construct preferences on the spot during the shopping process. Bettman et al. (1998) in particular discusses decision strategies that may be used in the selection process.

In this spirit, in this section, instead of assuming a specific optimization approach for the average subject, I use descriptive and unsupervised learning methods to examine the strategies that subjects pursued in participating in the experimental task. My interest is in trying to identify types of allocation approaches that subjects may have followed.

Table 7 summarizes a 7-cluster K-means analysis using subjects' percent allocated to funds according to their fee rank on the menu, the percent allocated based on RSI rank on the menu, and the number of funds selected; my interest here is to avoid using outcomes or pre-treatment characteristics in the clustering so as to focus clustering on subjects' menu choices. The clusters in Table 7 vary considerably in terms of pursuit of diversification and the extent to which fees or returns are pursued. Naturally, cluster analysis is an unsupervised learning technique that is used here for exploration. Figure 17 provides accompanying boxplots that summarize each clusters' distribution of feesProp, Number of funds selected, complexity, objective mutual fund knowledge (Scholl & Fontes (2022)), percent allocated to the lowest fee fund and percent allocated to the highest fee fund ***(complete summary statistics by cluster are in the appendix).***

Perhaps the most prominent result of Table 7 is the cluster I have labeled pure naive diversification. The average subjects in this cluster allocated almost precisely 20 percent to each of the five funds according to fee rank (18.137, 21.797, 21.794, 21.495, 16.777) and returns rank (20.883, 24.441, 19.931, 18.083, 16.662), and averaged about 4.5 funds. 3,350 subjects (48.5 percent) of the 6911 allocations that were made were identified in this cluster, making this cluster more than three and a half times the size of any other group. It seems that most subjects default into something approximating naive diversification. The median complexity level for this group is second highest (mean is 0.04), and the median mutual fund knowledge score is the second lowest across groups (mean is 3.04), and only about a third of participants identified in this group are owners of mutual funds.

The next group of note is labeled fee confused, with 386 subjects identified in this group, about 5.6 percent of overall subjects on the two survey administrations. This group diversified lightly with the median subject investing in only 2 funds (mean =2.44). While avoiding the problem of naive diversification, this group seems to have perplexingly targeted high fee funds, allocating an average of 72.9 percent to the highest fee fund. Oddly, this group also does not seem to have pursued a returns chasing strategy with 34.0 and 29.2 allocated respectively to the third and fourth ranked RSI funds, respectively. At the same time, in follow up survey questions, this group had similar responses to the pure naive diversification group in terms of a "you get what you pay for," heuristic, albeit with a slightly higher set of I don't know responses and a slightly lower number of respondents stating that in terms of mutual funds you get what you don't pay for. This group had the lowest average mutual fund literacy level of any group (2.94). Overall, even though a number of follow up questions were asked subsequent to the allocation task related to views on diversification and fee avoidance, it is hard to understand what these subjects were pursuing in the allocation task. One possibility is that they simply misunderstood the idea that higher fees mean lower returns. Another possibility identified in qualitative testing that was conducted as a precursor to this study, is that some individuals seem to

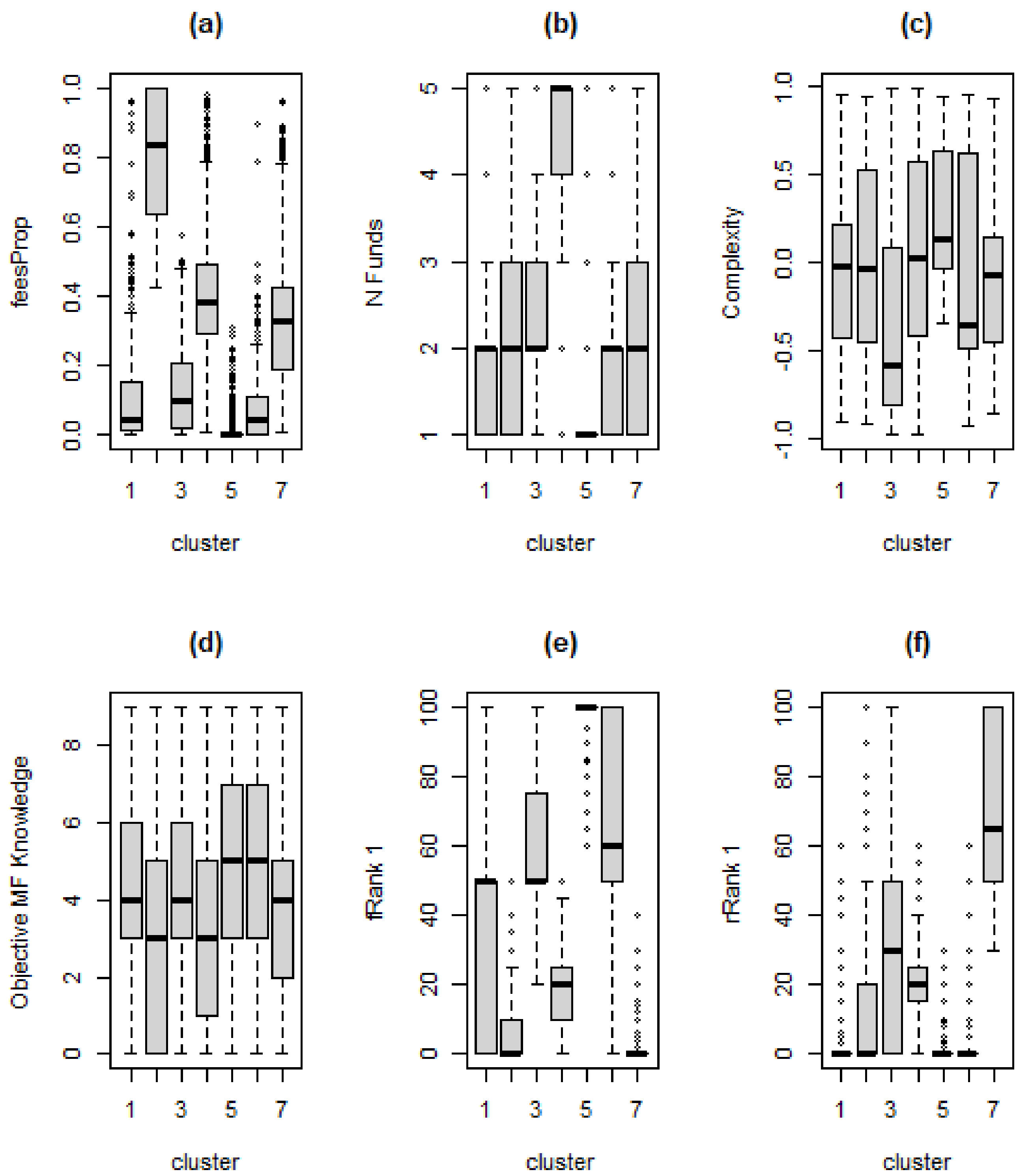


Figure 17: Boxplots of key elements of each cluster identified with 7-cluster K-Means.

actually target average values - so not high returns or low fees, but rather average values on some feature dimension or another. If taken seriously, that notion would suggest that subjects may have targeted middle-RSI funds, but tended to diversify slightly among the middle-RSI funds. In any event, one distinction between this group and the pure naive diversification group is clear: though much vilified in the literature and clearly a loss vis- a-vis the optimal allocation, naive diversification in the current choice context is actually a much better strategy than the strategy these subjects were pursuing; for low knowledge individuals, naive diversification may actually be a reasonable strategy to pursue.

The next cluster of note is the Optimal Strategy group. This group put 94.6 percent on average into the lowest fee fund. At the same time, note that they concentrated about the same fraction into the lowest RSI fund. The implication is that these subjects faced a relatively high degree of complexity, but in the face of such complexity, they pursued low fees rather than RSI. Indeed, the average complexity for this cluster was 0.277, about half a standard deviation higher than any other group. These subjects saw a complicated menu, saw through the RSI ruse, and went all in on the low fee fund. Perhaps not surprisingly, these subjects had the highest average mutual fund knowledge score at 5.07. This group also had higher levels of mutual fund ownership rates with 57.3 percent of participants owning funds, and higher levels of education that the other clusters, but these averages were largely in line with the values for other clusters.

The cluster labeled Returns Chasers, invested an average of 72 percent into the highest RSI fund, with some diversification on the menu (average of 2.2 funds allocated). This group ranked fifth in terms of mean mutual fund knowledge scores (3.82), fifth in ownership rates (45.8 percent), and had lower complexity levels than five other groups. They also tended to be older. About 13 percent of allocation subjects were classified as being in this cluster.

The remaining three clusters were somewhat similar to each other, in that they tended to perform reasonably well in terms of actual portfolio construction. The average feesProp for these three groups were 0.11, 0.13 and 0.084. They tended to be assigned modest to low levels of complexity (means of -0.04, -0.30, and 0.03), and the median participant in each cluster allocated to two funds. Their mutual fund knowledge scores were moderate, with averages ranging from 4.2-4.6 on the 9 point scale. The real distinction between the groups is presented in panels (e) and (f) of Figure 17. Groups 1 and 6, which had modest complexity levels, put little into the highest RSI funds. Instead, they put an average of 41 percent and 62 percent, respectively into the lowest fee fund, albeit with differential skewness in the two groups: the interquartile range for group 1 was about zero to fifty percent allocated to the lowest fee fund; while the interquartile range for cluster 6 ranged from about 50 to 100 percent in the lowest fee fund. Both of these clusters tended to put 80-90 percent of their allocation in the two lowest fee funds. Cluster 3 also put about 80 percent into the two lowest fee funds, but because they faced far lower than average levels of complexity, this is not easily distinguishable from a fee-chasing strategy.

	1	2	3	4	5	6	7
1	8	6	11	50	26	18	21
2	4	8	8	44	6	3	21
3	24	2	22	71	36	19	42
4	61	39	59	512	45	40	96
5	19	6	8	36	49	14	23
6	8	5	14	56	26	14	19
7	16	10	16	92	16	23	62

Table 8: Repeaters' strategy clusters in A1 (rows) and A2 (columns)

These clusters are only intended to be suggestive of strategies that subjects may have pursued in order to understand decisions they may have undertaken. Strategy stability is a somewhat different story. For subjects that took the experimental task twice, the issue of strategy stability arises. Did a participant pursue the same strategy in both A1 and A2? Table 8 provides a transition matrix between A1 cluster and A2 from the clusters identified above (for consistency within the rather unstable K-means, clusters were identified using the pooled observations).

Pure Naive Diversification not only was the most prevalent strategy in the experiments overall, but it was the most stable. 1834 subjects were in both A1 and A2, with 852 assigned to the Pure Naive Diversification cluster in A1. Of these, 512 (60 percent) pursued a Pure Naive Diversification strategy in A2. An additional 349 that were not Pure Naive

Diversifiers in A1 started pursuing this strategy in A2.

In A1, 155 of the repeater subjects pursued an Optimal Strategy in A1. For this group, the Optimal Strategy remained the dominant strategy in A2, with 49 subjects pursuing it again. Yet, this means that about two-thirds of subjects pursued something else in A2, meaning a worse allocation: 36 became pure naive diversifiers; 23 became return chasers; and 41 were assigned to clusters 1,3 and 6, which were not quite optimal, but also were relatively high performing strategies. Another 155 subjects opted in to the Optimal Strategy in A2 that had previously been in other strategies, more than half coming from clusters 1, 3 and 6.

Other strategy clusters showed far less consistency for the repeat subjects as evidenced by the diagonal of the transition matrix in Table 8. Clusters 1, 2, 3, 6 and 7 all had more subjects hop into naive diversification than remain in the strategy cluster. Returns Chasers (cluster 7) remained the second most popular strategy in A2 for subjects pursuing it in A1, but the diagonals are largely not favored on the table.

Taking a step back from the cluster analysis, Figure 18 highlights the number of funds chosen in A1 and A2. It illustrates the issue with a naive diversification strategy in that the typical subject that allocated to all five funds did poorly vs subjects that allocated to only one fund, but as discussed above, some of the participants that allocated to just a single fund did poorly. The diagram also highlights the fact that most subjects did not invest in a stable number of funds.

5.3 Mechanisms

Table 19 presents results of the interactions-augmented model on feesProp with potential theoretical mechanisms using the A1 experimental sample. I examine three potential mechanisms that may be at work: light salience; a you get what you don't pay for heuristic; and a strict salience approach following Bordalo et al. (2013) and related work. One immediate note about these results is that the inclusion of these additional terms does little to moderate the complexity effect, despite the inclusion of interaction terms with complexity. Light salience refers to the sense in which the term salience is most commonly used in the literature (for example, in Hartzmark & Sussman (2019)) - that is, if the presence or

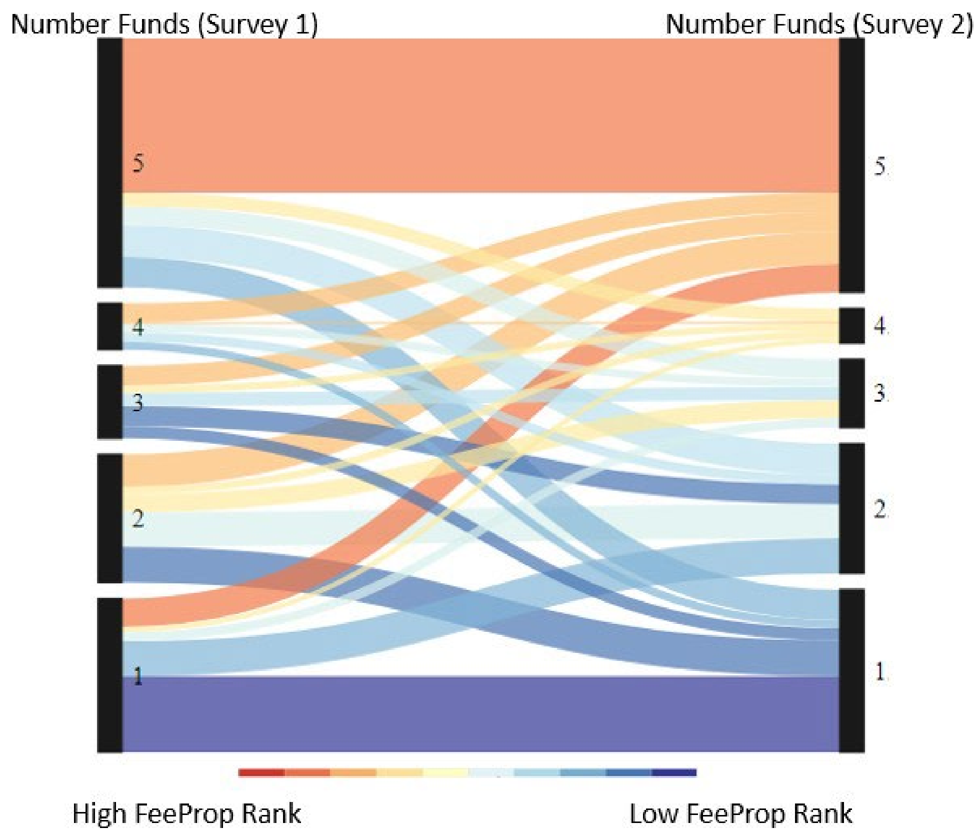


Figure 18: Transitions plot of number of funds allocated in A1 and A2. Color shading represents the average subjects' decile of feesProp Rank in the respective A1 (N) to A2 (N) grouping **based on the allocation in A1**. feesProp Rank summarizes the weighting towards high or low fee funds on the menu. Generally subjects that invested in all five funds did poorly, versus those that invested in just a single fund. However a portion of subjects invested in a single fund unwisely.

prominence of the information elements on the allocation screen affects allocation decision. This effect is identified off of placebo tests that dropped RSI and fee presentations from the allocation screen, although subjects still had access to the information in the summary prospectus. Column (1) presents this mechanism although this is the identical regression presented in the main results previously, except that the salience terms are now not suppressed. Results suggest that light salience does indeed have a role in the decision process, but it does not moderate the complexity effect.

Column (2) is the result of a heuristic I developed to understand subjects' inherent heuristic relating to the empirical relationship of fees and returns for mutual funds (a deeper dive into this heuristic is forthcoming in Scholl & VanEpps (2022)). The literature generally sees the correlation between fees and returns as negative: lower expenses typically means higher net returns. This need not be true in any real sense - after all, it could be that a law of one price exists in this market because market discipline penalizes low return funds' ability to charge fees. The question is clearly related to financial literacy, but I distinguish it because it is an empirical fact rather than some definition or concept that flows naturally from a traditional model. In fact, this empirical fact violates most assumptions about market discipline; Choi et al. (2009) and Hortaçsu & Syverson (2004) explicitly discuss the failure of the law of one price in the funds context. My focus here (see appendix for question text) asked subjects to respond if higher fees meant lower returns after fees.

Results suggest that only 17 percent of subjects had the you get what you don't pay for heuristic that aligns most closely with the empirical literature (note that 40 percent of respondents answered that they didn't know). The results here are suggestive, but by no means definitive, and merit some discussion. In column (2) the main effect of the protective heuristic (lower expenses leads to higher net returns) leads to an 8 percentage point reduction in fee levels. Although the coefficient on complexity is unchanged, the interaction suggests that the subjects that have this protective heuristic have a total magnitude of the complexity effect that is reduced by about two-thirds. The p-value on the interaction is less than 0.01.

Column (3) examines an alternative menu-effects mechanism: strict salience. Within the line of work of Bordalo et al. (2013) and their related series of papers, salience takes on an extremely strict form in that it is not the absence or prominence of certain features of information that gain attention, but how much the actual numerical pieces of information stand out within the menu. That is, price or quality that stands out relative to the other items on the menu will get attention. This is not easily mapped to the exact set-up of the current analysis, but a simple version is provided by including the standard deviation of fees and the standard deviation of RSI on the menu. Results suggest that inclusion of these salience terms may exacerbate investors' sensitivity to fees, but there is little evidence that the interactions explain away the complexity effect. The result suggests that salience may be important and complementary to the complexity effect.

	Dependent variable:		
	feesProp		
	(1)	(2)	(3)
complexity	0.08*** (0.01)	0.09*** (0.01)	0.07*** (0.01)
rsi salience	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
fee salience	0.08*** (0.01)	0.07*** (0.01)	0.07*** (0.01)
lower exp. = higher ret.		−0.08*** (0.01)	
sd(exp.)			−31.84*** (3.44)
sd(rsi)			−0.06*** (0.02)
complexity X rsi salience	0.07*** (0.02)	0.07*** (0.02)	0.07*** (0.02)
complexity X fee salience	0.05** (0.02)	0.05* (0.02)	0.05** (0.02)
complexity X lower exp. = higher ret.		−0.06*** (0.02)	
complexity X sd(exp.)			34.64*** (7.00)
complexity X sd(rsi)			0.08*** (0.03)
F Statistic	15.114***	16.412***	18.865***
Observations	4,021	4,021	4,021
R^2	0.14	0.16	0.18
Note:			*p<0.1; **p<0.05; ***p<0.01

Figure 19: Regressions of interactions-augmented models on feesProp with potential theoretical mechanisms on subjects in A1. Column (1) is the A1 baseline interactions-adjusted model discussed previously - here salience is identified off of placebo tests that dropped RSI or fee salience from the baseline presentation conditions (both were still provided in the summary prospectus mockup); column (2) augments the model with a survey question that asks subjects if they think the correlation between fees and returns is positive (you get what you pay for), negative or zero; columns (3)-(7) introduce stricter forms of salience in the sense of @bordalo2013salience and related work.

5.4 Recall and Learning

While a within-subjects approach may create learning from experience, it is unlikely to be the case in the current context. First, subjects were not provided with direct feedback on their performance in the experiment, so there is unlikely to be a learning process. Second, anecdotally, the survey administrators suggest that recall on this particular survey platform does not appear to be very strong. Finally, subjects were directly asked if they recalled similar questions (of which this experiment constituted a single question) being asked before - only a third of the subjects that took both experiments indicated any familiarity with the tasks and questions of the survey instrument.

Figure 20 provides a loess fit of complexity on feesProp. The top panel suggests that repeat subjects (those that participated in both A1 and A2) paid very slightly lower fees than non-repeat subjects (participated in A2, but not A1), yet the slope of the curves is little different. Moreover, after controlling for other characteristics in the bottom panel, repeat subjects fared no better than non-repeats. As suggested in Figure 21, repeat subjects fared no differently in the face of complexity in either experiment.

6 Welfare Implications

6.1 Experimental Loss

I calculate the investment loss attributable to the complexity effect as assigned within the experiment. This loss calculation is extremely conservative because the menu options employed in this paper consist of relatively low cost index funds, so that for some assigned menus, the differences in costs across the menu may be small relative to the differences on menus offered in say a brokerage investment account. Moreover, because the funds that I use are index funds, with only one true feature differentiating the funds, the welfare implications that I estimate should be considered a lower bound as compared with situations where a consumer may be comparing many different attribute features such as, in the investment context, risks, strategies, reputations, brands, etc. Nevertheless, the use of index funds provides the advantage of an unambiguous estimate on the welfare dimension - without consideration of any specific assumptions on preferences.

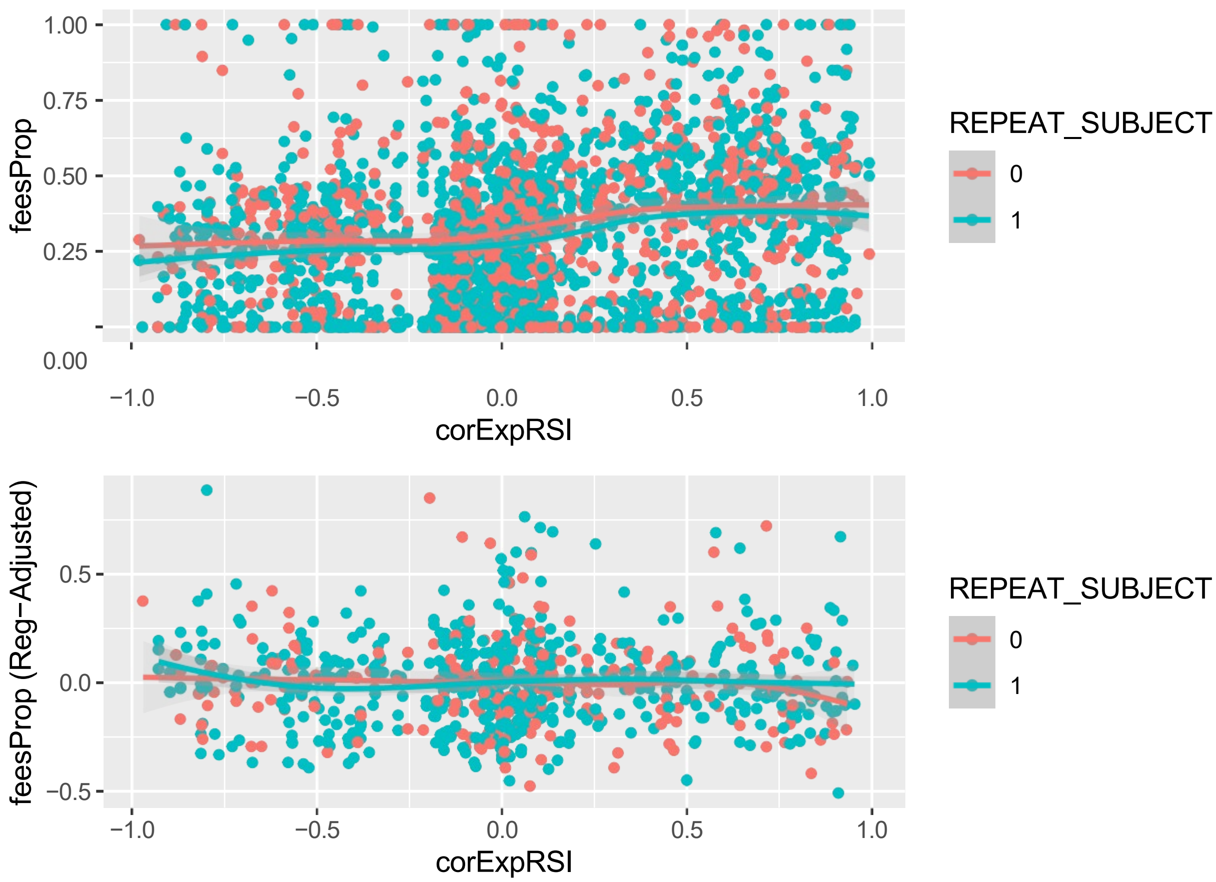


Figure 20: Learning for repeat subjects. Top panel: feesProp by complexity. Bottom panel: Complexity on feesProp after regression adjustment.

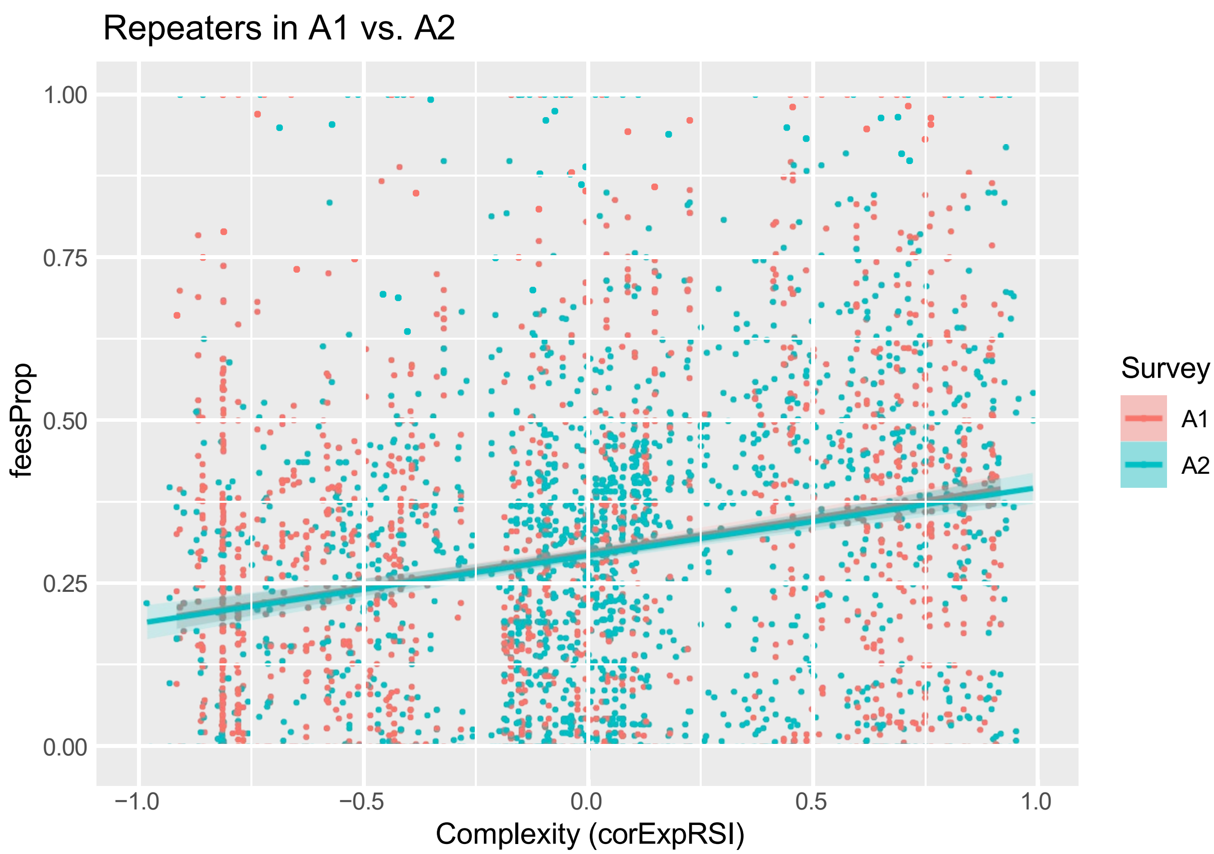


Figure 21: Graphical depiction of repeat subjects' portfolio allocation in response to complexity in A1 vs. A2.

To generate the estimate, I calculate two new counterfactual feesProp measures. The first is provided by $feesProp_i^{I} = max\{feesProp_i - \tau * (1 + \rho_i)), 0\}$, which eliminates the complexity effect across the complexity distribution ("deduction") with a lower bound of zero. As depicted in the top panel of Figure 22, this transform essentially constructs a new feesProp measure for each participant as if the were unaffected by complexity. I set $\tau = 0.09$ in line with the experimental results which ranged from 0.8-1.0. The bottom panel of 22 provides an estimate of the welfare loss associated with this effect alone - that is, the welfare loss if complexity had no effect. The second measure is $feesProp_i^{II} = min\{feesProp_i + \tau(2-(1+\rho)), 1\}$, which adds in a complexity effect across the complexity distribution ("add"). This second measure generates a counterfactual complexity effect for each participant as if they adopted their individual projected response to the maximum complexity assignment ($\tau(\rho = 1)$). The addition or subtraction is made with reference to the experimentally assigned complexity level, so that for subjects assigned to a complexity level of -1, no change in feesProp is observed in the first construct, but in the second construct, a subject that was assigned that value of complexity would have a feesProp measure that is higher by a value of 2τ; a subject assigned $\rho = 1$ would have values that are 2τ lower in the deduction counterfactual, and unadjusted in the second; and a subject receiving an intermediate complexity value would be shifted in both cases, but by a smaller amount in each case. On net, each of these counterfactuals for the individual subject leads to a flattening out of the regression line at one of the endpoints, with the difference in these two counterfactual lines representing the welfare loss. The wrinkle here is that rather than using the regression line, I do this for every subject separately so that I have an accurate accounting of the loss for each individual using their experimentally assigned menu.

To precisely formulate an estimate of the welfare loss specific to each subject's menu, I translate these counterfactual feeProp measures to an average expense ratio *on subject i's experimentally assigned menu k*: $AvgExpr_i^{k,*} = \frac{feesProp^{*} * (maxFees^k - minFees^k) + minFees^k}{10000}$, where the * denotes the respective counterfactual. Using this expense ratio, I then recursively calculate the actual fees paid and investment returns for a 100,000 dollar starting balance over a 25 year horizon using the assumed 13 percent annual return experienced during the most recent ten years (while mean welfare loss values are sensitive to the assumed average returns distribution, the shape of the distribution is not). The total (nominal) welfare loss is then the difference between the final investment balance using the two counterfactual average expenses.

The welfare loss is rather substantial even with this extremely conservative calculation. As highlighted in the bottom panel of Figure 22, the investment loss from the counter- factual generated by feesProp' alone averages about 26,000 dollars - about 2.5 times the original investment balance. A more complete picture is provided in Figure 23; the top panel presents each subject's final balance from the 25 year projection described above, the bottom panel is the histogram of total welfare loss. The black curve in the top panel represents the final balance using the aforementioned projection parameters based on subjects' actual experimental allocation on their experimentally assigned menu. The "deduct" curve is the resulting final balance from the counterfactual generated using feesProp', the "add" curve is generated from feesProp' ' - the total welfare loss depicted in the bottom panel is the difference in subjects' final balance projections under the two counterfactual feesProp measures. Again, welfare can be directly interpreted without consideration of specific utility functions because the choice options are unambiguously identical except for costs and their consumption value is already in dollar terms. Note that "add" reduces welfare because it refers to adding in the maximal complexity effect for subjects that had not been exposed to complexity. The total welfare loss has a median of 58,937 dollars, a mean of 63,496 dollars, the 75th percentile is 94,518 dollars and the maximum is 105,890 dollars. Again, these are nominal dollars, but deflating with a 2.5 percent annual inflation rate assumption renders the mean as 34,249 dollars and the 75th percentile as 50,982 dollars. The aggregate loss for just these 6,911 participants on the experiment is then 438,821,520 dollars (236,696,198 in deflated dollars), which seems substantial for so few participants. Applying survey weights to the A1 sample to provide a population estimate yields a slightly lower nominal mean of 59,502 dollars, or 32,094 in fixed current dollars.

Projection of this conservative welfare estimate into the population as a whole is somewhat more complicated. A thorough assessment requires individual investment levels, menu options, menu choices, and so forth. I do not have access to a data source that would provide such complete details for a representative sample of the population. Nevertheless, a cursory projection may provide some insight on the magnitude. Due to the probability- based sampling frame of the survey instrument on which the experiment was deployed, subjects that participated in the experiment are fairly representative of the population as a whole. As such, it is not unreasonable to assume that the effect sizes for the experimental sample are representative of the effect sizes in the population; that is, were one to entreat the entire population to perform the experimental allocation task, they would fare similarly. As such, applying the 32,094 value to the prime working age population of the United States [11] who may be making retirement decisions representative of the current set-up yields an estimate of 4.159 trillion dollars. Interpretation of this cost estimate must be made carefully: this estimate is not the actual cost in the population, but rather the projected inflation-adjusted consequences over a 25 year horizon were the population to engage in the experimental task of the main experiment. On one hand, as with the cost consequences estimated for the experimental subjects above, this is likely to be too conservative vis-a-vis actual menus that individuals are faced with in the marketplace, and the consequences of complexity for multiple feature dimensions.

On the other hand, this estimate is very much an estimate of the costs within the bounds of the experiment; in the actual marketplace, individuals may have far lower investment balances, lower investment participation rates, and so forth. To refine this estimate somewhat, using the 2019 Survey of Consumer Finances (SCF), I constructed a measure of investible assets, comprised of directly held pooled investment funds, IRA balances and employment related investment accounts; this total investment balance represents largely mutual fund investments that present a similar decisionmaking context to the choice setup of this paper. From the SCF data, I constructed mean, median, and population-weighted size estimates for each cohort year of individual age and used these values to construct a projection of the welfare loss for each age cohort using the methods above. Here, I replaced the initial investment balance (previously 100,000 dollars) with the mean or median balance of investable assets in a given age cohort, and projected forward a number of years equal to 65 less the respective age group applying the average expense ratio from the "add" and "deduct" methods above from the experimental estimates. This approach gives me a

[11]Ages 25-54. Projections made using April 2020 population estimates from census.

snapshot estimate of the current total welfare loss in each age group based on average holdings within the cohort group and a retirement projection horizon for that age group (here I mean current in the sense that it does not make a lifecycle assumption for individuals so that 18-year olds do not accumulate more assets over time) . I then multiply this individual estimate of the average welfare loss in each age group by the estimated population in each group. Using the SCF median for each group as the base investment amount, the welfare loss estimate using this approach is 144.3 billion dollars, using the SCF's (heavily skewed) mean, the estimate is 3.37 trillion dollars - 1.76 trillion in deflated dollars.[12] These estimates are somewhat more tenuous than the previous estimates, but indicative; they are surely somewhat overstated in terms of the true incidence of the complexity cost in the population as that they assume each individual with a funds-like investment was subjected to a high complexity menu, but they are again understated in the sense that average menu costs and total complexity from the index fund menu in the experimental sample is likely to be far lower than what these consumers actually faced in making their investment choices.

[12]The deflator here deflates each cohort by the commensurate number of years forward in which the cohort's projection is made. For example, the 65 year old cohort is not deflated, while the 18 year old is deflated by 1.025ˆ47.

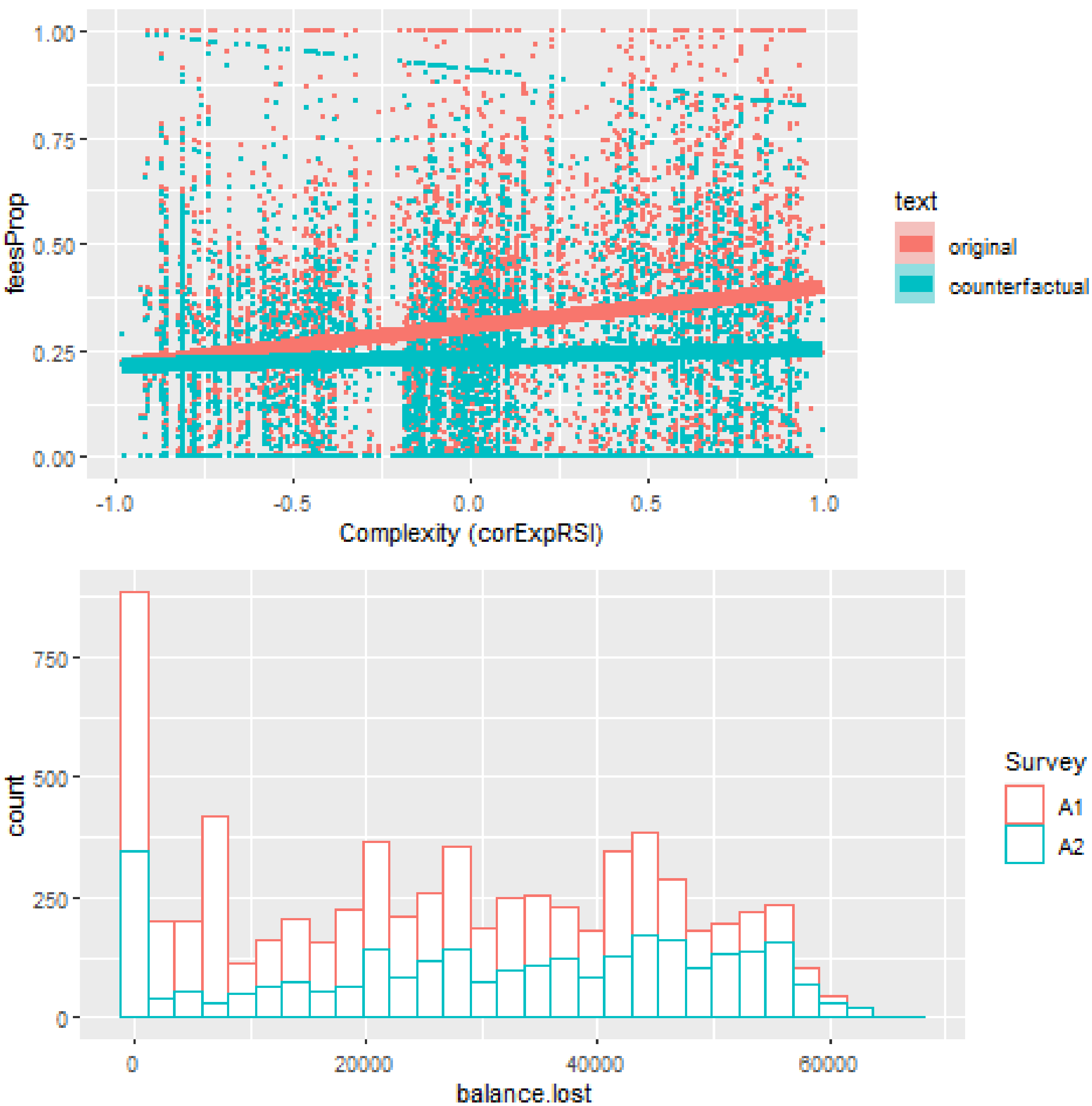


Figure 22: Welfare implications of choices made by study participants in A1 and A2.

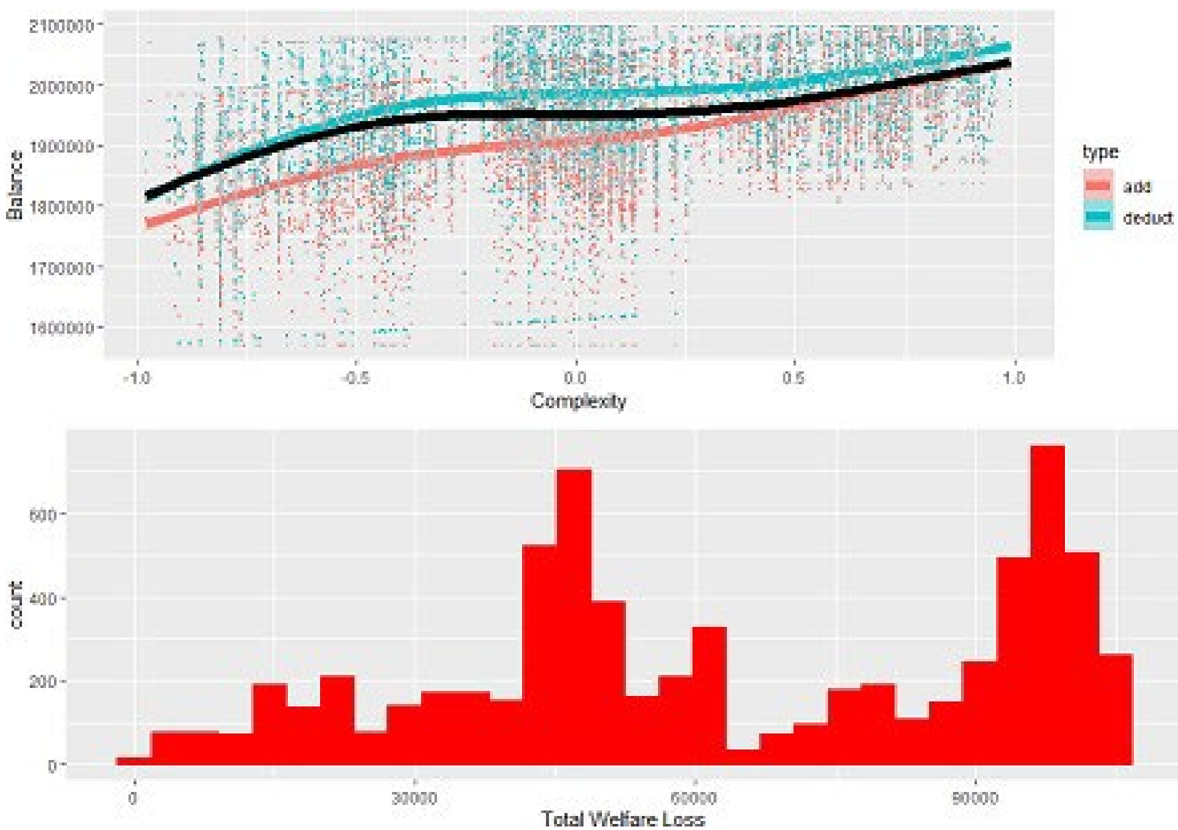


Figure 23: The black line represents the 25 year projected final balance using subjects' chosen allocation on their assigned menu using a 13 percent average return. Total welfare loss due to complexity effect is illustrated by the difference between the red and the blue fits to the respective counterfactual final investment balance as described in the text.

7 Conclusion

This paper develops the concept of choice-set complexity and uses an experimental framework to examine its consequences for consumer choice. Using a case study of investment decisionmaking, in my experimental setting, subjects faced with higher complexity make significantly worse allocation choices - an increase of a normalized sixteen to twenty percentage points in menu-normalized costs for a move from my lowest to highest states of complexity. Results appear to hold across subgroups even though the levels of fees paid do differ by subgroup. A placebo test and a replication experiment lend credence to the findings. Results hold for both between and within subjects estimates, and are largely unaffected by estimation.

The randomization inference framework is conservative and precise, but ultimately of limited value. The concern about breech of the randomization protocol, covariate imbalance and incidental correlation between treatment and observables seems to have had virtually no consequences for the results, with the fact that the menus were randomly assigned seeming to override the specifics of the random assignment mechanism. Although not discussed in the paper, it is noteworthy that bias vis-a-vis estimates where covariates are introduced without the Lin method was fairly small, although I did encounter interaction terms that were problematic when the proper adjustment was not performed.

The interest in decision strategies suggests a few exploratory clusters of strategy groupings that could characterize how individuals pursue the decisionmaking task. In the investment case study, pure naive diversification is both most prevalent and most stable, but it is somewhat better than a returns chasing strategy in a highly complex context. What should be eye-opening to those that view the choice context as straightforward, few pursued the optimal strategy, and those that did not pursue it had very severe welfare losses. Most other subjects pursued some lighter form of naive diversification, but seemed to chase one feature or another based on the overall menu complexity and other factors.

Without a utility function assumption, I am still able to provide plausible welfare estimates for this particular case study. The results are suggestive that complexity of the choice set is an important obstacle to optimal decisionmaking that may create deviations from optimal welfare in many different consumer contexts. In the specific investment case study that I pursue, the welfare loss from choice set complexity is substantial: resulting in a projected welfare loss for my experimental subjects of some USD 430 million over a reasonable investment horizon. At the same time, choice set complexity is not merely an artifact of my experimental framework - analysis of administrative records from employee-sponsored retirement plans suggests that choice set complexity is widely prevalent in such plans.

Policy measures to mitigate the consequences of choice set complexity are perhaps worthy of consideration. At the individual decisionmaker level, education and awareness might be an effective place to start. Yet, while high sophistication individuals - those with at least 7 of 9 questions correct in the abridged Scholl & Fontes (2022) index - seem to have been considerably less susceptible to complexity, they were still affected to a certain degree. Moreover, a mere 6.5 percent of participants in A1 scored this high. Another option might be to focus on reprogramming individual beliefs about the relationship between quality and price; subjects that held the correct belief about the empirical relationship, seem to be inoculated to the complexity effect. Of course, this belief is particularly advantageous in the current experimental set-up, and it seems plausible that one could curate alternative menus of financial products in which this belief leads to sub-optimal allocations, and in other cases price and quality may be positively correlated.

Alternatively, prescriptive nudges that induce consumers to adopt an appropriate decision strategies might help investors to formulate strategies that protect against complexity in the decisioning environment, but these too seem difficult to implement; real-world variation of menus and features as well as their persistent evolution makes it difficult to imagine a choice strategy prescription that would keep individuals ahead of the evolution in menus (indeed, in many contexts it is well known that investors are extremely inert in their decisionmaking Beshears et al. (2009)). Moreover, even though the context that I study has a unique strictly dominant strategy that is invariant to complexity, a slight relaxation of the context to broaden the choice set - for example to different types of indexed mutual funds - already makes the decision process more complex and the optimal decision strategy less straightforward to prescribe. Furthermore, it cannot be overlooked that even at the lowest levels of complexity the average subject's allocation is about 20 percentage points on the menu-normalized scale - about as large as the full complexity effect itself. This level estimate of the loss in the decision problem can be thought of as a pure baseline loss due to a lack of knowledge and sophistication, inexperience with investment decision-making, the baseline complexity of investment decisionmaking and so forth. In addition, based on qualitative interviews, it is possible that the investment setting itself is inherently stressful and emotionally costly to subjects in the first place and may push subjects into emotion-driven rather than optimizing strategies (e.g. Loewenstein & Lerner (2003)) - to be clear, my design in this paper does not specifically allow me to identify an emotional avoidance channel. In any event, at least in the investment case-study, it seems that the choice environment may simply be too complicated and that hopes that education will be an effective solution seem ephemeral - leading to a baseline welfare loss even in the best of choice set complexity circumstances. The current study is perhaps not sufficiently generalized to consider recommendations in other choice domains where consumers may have adverse welfare implications of complexity.

Yet, continuing with the investment case study, targeting education to plan benefit administrators in firms' human resources offices might be a more effective means of mitigating the consequences. After all, it is generally these individuals within a company that establish relationships with plan administrators, effectively determining the extent of choice set complexity that all employees in the firm are faced with. Targeted education policies or other efforts to eliminate the choice complexity generated by menus in the first place might avert placing unsophisticated employees in difficult choice environments in which they are unlikely to succeed.

8 Appendix

8.1 Identification of Candidate Funds

To generate complexity on individual menus as per the example in Section 3, I identified twelve candidate S&P 500 index funds so that the resulting set of 792 (12 choose 5) portfolio permutations provided a covering of the space on the principal measure of menu complexity. The twelve candidate funds were sampled from the universe of active S&P500 index funds that had 10 years of operating history and were active in the 2014-2016 period (the historical range was necessary in order to provide summary prospectus documents that had comparable histories). Candidates were identified by a mix of analysis and inspection.

It is important to note that while the effort was designed to select a parsimonious set of funds, the objective was to get a broad distribution of the properties of resulting menus while limiting the total number of funds selected. After taking an initial sample for computational tractability, I then examined the distribution of resulting portfolios generated, with the primary variable of interest the intra-menu correlation between RSI and the ex- pense ratio. I then identified which individual funds appeared in portfolio sets with the desired properties - in particular, I identified funds that might appear on both high com- plexity menus and low complexity menus, with the goal of narrowing the field to about 10-12 funds [13]. After identifying an appropriate number of funds with the desired proper- ties (ultimately the 12 candidates), I simulated the properties of the full set of 12 chose 5 menus resulting from this collection of funds to verify that a broader set of properties was satisfied (details are available in the online appendix). Such features included the ratio of the maximum and minimum fee on a menu, the standard deviation of fees and returns on a menu, average fees and so forth. In order to obtain a set for which resulting menus would have the desired properties, this process was iterated a few times in order to consider various options.

[13] To simply take portfolios with the desired complexity properties (say low and high intra-menu correlation), would result in a set of portfolios that ideally matched the desired properties of the manipulation variable, but potentially each had a different set of funds. For operational reasons (experiment programming as well as mock-ups of summary prospectuses and other features), a target range of 10-12 funds were desired.

8.2 Prospectus Mock-Ups

Figure 24 provides a full summary prospectus for one of the twelve candidate funds. All summary prospectus documents were mocked up from a common template. The highlighted text indicates text that was changed across the funds.

8.3 Completion Rates by Experimental Condition

Groups vary in size and composition, but visible inspection of this table does not provide an easy test of the completion and qualification rates. To examine this directly, I conducted group-wise 2-sample t-tests of differences in means (unequal variances) for each of the three rate statistics. The p-values from these tests indicate rejection of the null hypothesis of equality in 5 percent, 15 percent and 6 percent of cases, for each rate variable, respectively. While this suggests some slightly higher difference in experimental completion rates among the groups above what would be expected by chance alone, the null is rejected in only 15 percent of the 212 cases and does not appear to be a cause for great concern.

8.4 Heuristics Questions Text

Which of the following best describes your view on the relationship between a fund's expenses and performance:

[RANDOMIZE OPTIONS 1-3] 1. Funds with higher expenses (i.e. fees) tend to have higher net returns (after expenses) 2. Funds with lower expenses (i.e. fees) tend to have higher net returns (after expenses) 3. There is no relationship between expenses (i.e. fees) and returns 77. I don't know

HYDE® 500 Index Fund

Summary Prospectus
June 1, 2017

HYDE

Investments

Fund Summary

HYDE® 500 Index Fund

Investment Objective

The fund seeks to provide investment results that correspond to the total return (*i.e.,* the combination of capital changes and income) performance of common stocks publicly traded in the United States.

Fee Table

The following table describes the fees and expenses that may be incurred when you buy and hold shares of the fund.

Annual Operating Expenses
(expenses that you pay each year as a % of the value of your investment)

Management Fee	0.20%
Other Expense	1.00%
Fee Waiver	(0.00%)
Total Annual Operating Expenses	1.20%

Example

The Example is intended to help you compare the cost of investing in the fund with the cost of investing in other mutual funds. The Example assumes that you invest $10,000 in the fund for the time periods indicated and then hold or redeem all of your shares at the end of those periods. The Example also assumes that your investment has a 5% return each year and that the fund's operating expenses remain the same. Although your actual costs may be higher or lower, based on these assumptions your costs would be:

1 Year	3 Years	5 Years	10 Years
$126	$397	$696	$1,585

Portfolio Turnover

The fund pays transaction costs, such as commissions, when it buys and sells securities (or "turns over" its portfolio). A higher portfolio turnover rate may indicate higher transaction costs and may result in higher taxes when fund shares are held in a taxable account. These costs, which are not reflected in annual operating expenses or in the example, affect the fund's performance. During the most recent fiscal year, the fund's portfolio turnover rate was 5% of the average value of its portfolio.

Principal Investment Strategies

Figure 24: Prospectus mock-up for the fund anonymized as Hyde. All prospectuses mocked up from a common template. Highlighted te8x0t indicates fields that have been changed.

Table 9:

condition	N	pre comp rate	SE	comp rate	SE	Exp comp rate	SE
1	269	0.9442379	0.0139905	0.7598425	0.0260456	0.7583643	0.0261002
2	273	0.9632353	0.0113894	0.7718631	0.0253972	0.7838828	0.0249109
3	261	0.9501916	0.0134659	0.7773279	0.0257522	0.7854406	0.0254103
4	267	0.9396226	0.0145767	0.7510040	0.0264644	0.7340824	0.0270390
5	260	0.9536680	0.0130362	0.7560976	0.0266324	0.7346154	0.0273830
6	237	0.9535865	0.0136656	0.7399103	0.0284955	0.7257384	0.0289800
7	249	0.9274194	0.0164418	0.7105263	0.0287405	0.6867470	0.0293932
8	259	0.9147287	0.0173539	0.7245763	0.0277583	0.6988417	0.0285060
9	147	0.9315068	0.0208333	0.8029197	0.0328095	0.7823129	0.0340367
10	259	0.9571984	0.0125771	0.7822581	0.0256446	0.7992278	0.0248907
11	344	0.9501466	0.0117345	0.8012232	0.0215169	0.7994186	0.0215900
12	276	0.9418182	0.0140904	0.7704280	0.0253146	0.7681159	0.0254035
13	237	0.9322034	0.0163300	0.7227273	0.0290781	0.7215190	0.0291171
14	261	0.9498069	0.0135151	0.7845528	0.0254484	0.7969349	0.0249005
15	264	0.9272031	0.0159898	0.7591837	0.0263157	0.7386364	0.0270418
16	246	0.9387755	0.0152854	0.6869565	0.0295665	0.6951220	0.0293512
17	264	0.9541985	0.0128664	0.7609562	0.0262492	0.7727273	0.0257920
18	230	0.9478261	0.0146631	0.7314815	0.0292230	0.7347826	0.0291083
19	250	0.9277108	0.0163785	0.7575758	0.0271039	0.7400000	0.0277417
20	159	0.9050633	0.0232465	0.7569444	0.0340162	0.7232704	0.0354797
21	275	0.9780220	0.0088410	0.7798507	0.0249861	0.8290909	0.0226996
22	365	0.9311295	0.0132549	0.7624633	0.0222755	0.7534247	0.0225605

Table 10: Summary Statistics

CLUSTER		1			2			3			4			5			6			7		
Variable	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	Test
OWNER	445	0.548	0.498	319	0.263	0.441	551	0.546	0.498	2827	0.348	0.476	534	0.573	0.495	422	0.573	0.495	757	0.458	0.499	F= 43.847***
AGE	518	50.62	15.593	386	47.119	16.242	626	49.331	15.507	3350	45.727	16.206	643	49.333	15.989	479	48.987	15.369	909	50.146	16.029	F= 17.69***
hispanic	518	0.1	0.301	386	0.184	0.388	626	0.096	0.295	3350	0.164	0.37	643	0.081	0.273	479	0.094	0.292	909	0.125	0.331	F= 11.571***
retired	518	0.191	0.394	386	0.145	0.353	626	0.179	0.384	3350	0.144	0.351	643	0.163	0.37	479	0.154	0.362	909	0.168	0.374	F= 2.154**
EDUC	518	11.239	1.61	386	10.521	1.639	626	11.385	1.553	3350	10.681	1.651	643	11.527	1.688	479	11.42	1.573	909	11.146	1.578	F= 52.107***
mfknow_obj_score	445	4.389	2.474	319	2.944	2.306	551	4.232	2.441	2827	3.049	2.252	534	5.073	2.588	422	4.685	2.599	757	3.815	2.323	F= 94.853***
corExpRSI	518	-0.043	0.466	386	-0.008	0.563	626	-0.309	0.572	3350	0.046	0.537	643	0.277	0.354	479	-0.031	0.57	909	-0.084	0.447	F= 78.358***
numfunds	518	2.046	1.141	386	2.448	1.494	626	2.492	1.293	3350	4.441	1.051	643	1.348	0.777	479	1.804	1.039	909	2.215	1.262	F= 1387.427***
YGWYPF	445	0.133	0.34	319	0.169	0.376	551	0.109	0.312	2827	0.178	0.383	534	0.096	0.294	422	0.116	0.321	757	0.155	0.362	F= 7.055***
YGWYDONTPF	445	0.218	0.413	319	0.116	0.321	551	0.194	0.396	2827	0.148	0.355	534	0.303	0.46	422	0.265	0.442	757	0.161	0.368	F= 18.902***
YGWYPF_norelation	445	0.344	0.476	319	0.223	0.417	551	0.338	0.473	2827	0.224	0.417	534	0.318	0.466	422	0.334	0.472	757	0.317	0.466	F= 13.884***
YGWYPF_IDK	445	0.288	0.453	319	0.489	0.501	551	0.356	0.479	2827	0.432	0.495	534	0.279	0.449	422	0.275	0.447	757	0.357	0.479	F= 18.708***
sdexpr	518	0.003	0.001	386	0.003	0.001	626	0.003	0.001	3350	0.003	0.001	643	0.003	0.001	479	0.003	0.001	909	0.004	0.001	F= 17.36***
sdRSI	518	1.148	0.292	386	1.204	0.287	626	1.18	0.242	3350	1.19	0.286	643	1.3	0.283	479	1.109	0.265	909	1.225	0.292	F= 26.962***
avgexpr	518	0.004	0.001	386	0.004	0.001	626	0.005	0.001	3350	0.004	0.001	643	0.004	0.001	479	0.004	0.001	909	0.005	0.001	F= 76.228***
avgRSI	518	7.47	0.383	386	7.43	0.384	626	7.555	0.414	3350	7.45	0.415	643	7.33	0.403	479	7.62	0.348	909	7.456	0.411	F= 30.141***
feesProp_RANK	518	183.969	68.351	386	438.291	60.21	626	167.104	50.363	3350	296.978	43.371	643	108.555	18.61	479	154.427	62.242	909	289.153	61.375	F= 3079.506***
invfeerank1	518	41.954	33.952	386	5.426	11.013	626	62.267	21.716	3350	18.137	10.594	643	94.608	10.511	479	62.219	34.155	909	3.502	7.429	F= 3024.305***
invfeerank2	518	41.376	33.05	386	6.008	11.868	626	19.896	20.425	3350	21.797	14.729	643	3.323	7.969	479	27.434	31.916	909	30.074	34.346	F= 213.187***
invreturnsrank1	518	5.124	11.712	386	15.863	27.972	626	33.324	30.745	3350	20.883	11.613	643	1.411	5.121	479	6.534	14.128	909	72.09	22.636	F= 1658.916***
feesProp	518	0.114	0.164	386	0.81	0.182	626	0.133	0.131	3350	0.394	0.166	643	0.015	0.043	479	0.084	0.128	909	0.34	0.199	F= 1592.466***
RISKaversion	426	2.901	0.884	285	2.909	0.867	499	2.9	0.844	2629	2.942	0.846	513	2.899	0.818	395	2.934	0.849	741	2.883	0.86	F= 0.7
important_to_spread	410	3.71	1.088	274	3.646	1.133	486	3.628	1.103	2555	3.659	1.096	502	3.649	1.065	380	3.626	1.105	719	3.658	1.08	F= 0.278
returnsimportant	411	4.073	0.905	274	3.96	0.962	482	4.037	0.845	2546	4.039	0.913	499	4.118	0.857	380	4.095	0.839	718	4.11	0.839	F= 1.716
feesimportant	413	4.245	0.859	272	4.202	0.855	489	4.249	0.781	2558	4.237	0.82	503	4.28	0.798	384	4.227	0.743	720	4.265	0.731	F= 0.45
returnsareskill	387	3.53	0.861	260	3.55	0.901	457	3.451	0.919	2429	3.523	0.899	478	3.552	0.854	366	3.577	0.8	686	3.507	0.889	F= 0.896

Statistical significance markers: * p¡0.1; ** p¡0.05; *** p¡0.01

Table 11: Summary Statistics

CLUSTER		1			2			3			4			5			6			7		
Variable	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	Test
OWNER	305	0.541	0.499	243	0.243	0.43	413	0.54	0.499	1966	0.337	0.473	330	0.594	0.492	291	0.57	0.496	473	0.448	0.498	F= 34.969***
AGE	305	49.377	15.888	243	46.066	16.062	413	48.913	15.317	1966	44.924	16.192	330	48.848	15.734	291	48.842	15.041	473	49.222	15.515	F= 10.765***
hispanic	305	0.102	0.303	243	0.169	0.375	413	0.09	0.286	1966	0.156	0.363	330	0.079	0.27	291	0.086	0.281	473	0.108	0.31	F= 6.417***
retired	305	0.167	0.374	243	0.128	0.334	413	0.165	0.371	1966	0.132	0.338	330	0.167	0.373	291	0.158	0.365	473	0.161	0.368	F= 1.404
EDUC	305	11.243	1.612	243	10.626	1.578	413	11.412	1.589	1966	10.686	1.657	330	11.518	1.51	291	11.512	1.552	473	11.252	1.494	F= 33.055***
mfknow_obj_score	305	4.331	2.472	243	2.811	2.354	413	4.254	2.485	1966	3.004	2.237	330	5.161	2.534	291	4.746	2.693	473	3.6	2.255	F= 71.64***
corExpRSI	305	-0.079	0.503	243	-0.033	0.607	413	-0.475	0.525	1966	0.02	0.596	330	0.349	0.369	291	-0.042	0.573	473	-0.15	0.485	F= 77.074***
numfunds	305	2.066	1.168	243	2.539	1.511	413	2.363	1.288	1966	4.429	1.039	330	1.367	0.848	291	1.811	1.055	473	2.279	1.29	F= 761.878***
YGWYPF	305	0.134	0.342	243	0.193	0.396	413	0.114	0.318	1966	0.174	0.379	330	0.076	0.265	291	0.12	0.326	473	0.161	0.368	F= 5.593***
YGWYDONTPF	305	0.21	0.408	243	0.091	0.288	413	0.201	0.401	1966	0.146	0.354	330	0.33	0.471	291	0.261	0.44	473	0.15	0.358	F= 16.803***
YGWYPF_norelation	305	0.338	0.474	243	0.21	0.408	413	0.334	0.472	1966	0.22	0.415	330	0.303	0.46	291	0.326	0.47	473	0.315	0.465	F= 9.527***
YGWYPF_IDK	305	0.302	0.46	243	0.502	0.501	413	0.349	0.477	1966	0.439	0.496	330	0.288	0.453	291	0.285	0.452	473	0.37	0.483	F= 12.638***
sdexpr	305	0.003	0.001	243	0.003	0.001	413	0.003	0.001	1966	0.003	0.001	330	0.003	0.001	291	0.003	0.001	473	0.003	0.001	F= 13.802***
sdRSI	305	1.118	0.272	243	1.196	0.27	413	1.164	0.198	1966	1.175	0.268	330	1.319	0.269	291	1.114	0.256	473	1.212	0.274	F= 23.037***
avgexpr	305	0.004	0.001	243	0.004	0.002	413	0.006	0.001	1966	0.004	0.002	330	0.004	0.001	291	0.004	0.001	473	0.005	0.001	F= 76.341***
avgRSI	305	7.533	0.357	243	7.464	0.363	413	7.621	0.372	1966	7.508	0.387	330	7.369	0.385	291	7.627	0.34	473	7.533	0.373	F= 19.081***
feesProp_RANK	305	185.243	71.828	243	434.895	60.661	413	159.099	51.388	1966	295.338	45.056	330	108.312	17.631	291	153.833	64.307	473	281.457	61.787	F= 1682.216***
invfeerank1	305	43.351	34.43	243	6.08	11.543	413	65.845	23.055	1966	18.486	10.976	330	94.524	10.772	291	64.357	33.623	473	4.133	8.017	F= 1610.442***
invfeerank2	305	37.957	32.258	243	5.914	11.343	413	19.838	20.394	1966	21.974	14.8	330	3.455	8.224	291	24.368	30.325	473	34.018	35.302	F= 119.377***
invreturnsrank1	305	5.485	12.432	243	18.091	29.897	413	36.073	32.072	1966	20.949	12.033	330	1.312	5.05	291	6.289	13.947	473	71.089	22.669	F= 785.118***
feesProp	305	0.12	0.172	243	0.8	0.186	413	0.116	0.132	1966	0.391	0.174	330	0.014	0.04	291	0.087	0.131	473	0.33	0.209	F= 857.787***
RISKaversion	252	2.917	0.9	191	2.984	0.861	333	2.934	0.841	1579	2.942	0.849	274	2.832	0.826	251	2.976	0.834	390	2.844	0.848	F= 1.558
important_to_spread	241	3.734	1.067	181	3.641	1.1	325	3.569	1.13	1540	3.656	1.096	269	3.647	1.043	243	3.564	1.102	376	3.686	1.06	F= 0.87
returnsimportant	243	4.123	0.91	181	3.945	0.987	323	4.031	0.856	1530	4.044	0.917	271	4.048	0.883	244	4.07	0.851	377	4.077	0.867	F= 0.791
feesimportant	243	4.288	0.828	181	4.193	0.864	326	4.242	0.804	1543	4.239	0.823	272	4.239	0.814	245	4.204	0.768	379	4.216	0.742	F= 0.359
returnsareskill	231	3.532	0.858	173	3.52	0.912	306	3.438	0.933	1455	3.533	0.9	257	3.514	0.848	230	3.557	0.811	359	3.479	0.874	F= 0.685

Statistical significance markers: * p¡0.1; ** p¡0.05; *** p¡0.01

References

Abadie, A., Athey, S., Imbens, G. W. & Wooldridge, J. M. (2014), Finite population causal standard errors, Technical report, National Bureau of Economic Research.

Abeler, J. & Jäger, S. (2013), 'Complex tax incentives-an experimental investigation'.

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

Angrist, J. D. & Pischke, J.-S. (2008), *Mostly harmless econometrics: An empiricist's companion*, Princeton university press.

Athey, S. & Imbens, G. W. (2017), The econometrics of randomized experiments, *in* 'Handbook of economic field experiments', Vol. 1, Elsevier, pp. 73–140.

Benartzi, S. & Thaler, R. H. (2001), 'Naive diversification strategies in defined contribution saving plans', *American economic review* **91**(1), 79–98.

Besedeš, T., Deck, C., Sarangi, S. & Shor, M. (2012), 'Age effects and heuristics in decision making', *Review of Economics and Statistics* **94**(2), 580–595.

Beshears, Choi, L. M. (2011), 'How does simplified disclosure affect individuals' mutual fund choices?', *Explorations in the Economics of Aging* pp. 75–96.

Beshears, J., Choi, J. J., Laibson, D. & Madrian, B. C. (2009), 'The importance of default options for retirement saving outcomes', *Social security policy in a changing environment* pp. 167–195.

Beshears, J., Choi, J. J., Laibson, D. & Madrian, B. C. (2011), How does simplified disclosure affect individuals' mutual fund choices?, *in* 'Explorations in the Economics of Aging', University of Chicago Press, pp. 75–96.

Beshears, J., Choi, J. J., Laibson, D. & Madrian, B. C. (2013), 'Simplification and saving', *Journal of economic behavior & organization* **95**, 130–145.

Bettman, J. R., Johnson, E. J., Luce, M. F. & Payne, J. W. (1993), 'Correlation, conflict, and choice.', *Journal of Experimental Psychology: Learning, Memory, and Cognition* **19**(4), 931.

Bettman, J. R., Luce, M. F. & Payne, J. W. (1998), 'Constructive consumer choice processes', *Journal of consumer research* **25**(3), 187–217.

Bettman, J. R., Luce, M. F. & Payne, J. W. (2008), 'Consumer decision making: A choice goals approach.'.

Bordalo, P., Gennaioli, N. & Shleifer, A. (2013), 'Salience and consumer choice', *Journal of Political Economy* **121**(5), 803–843.

Brocas, I., Carrillo, J. D., Combs, T. D. & Kodaverdian, N. (2015), 'Consistency in simple vs. complex choices over the life cycle'.

Carvalho, L. & Silverman, D. (2019), Complexity and sophistication, Working Paper 26036, National Bureau of Economic Research.
URL: *https://www.nber.org/papers/w26036*

Chin, A., Scholl, B. & VanEpps, E. (2021), 'Jargon in fund fee disclosures', *Working Paper, SEC Office of the Investor Advocate* .

Choi, J. J., Laibson, D. & Madrian, B. C. (2009), 'Why Does the Law of One Price Fail? An Experiment on Index Mutual Funds', *The Review of Financial Studies* **23**(4), 1405–1432.
URL: *https://doi.org/10.1093/rfs/hhp097*

Deaton, A. (1997), *The analysis of household surveys: a microeconometric approach to development policy*, The World Bank.

Dellaert, B. G., Johnson, E. J. & Baker, T. (2019), 'Choice architecture for healthier insurance choices: Ordering and partitioning can improve decisions', *ERIM Report Series Reference* .

DellaVigna, S. (2009), 'Psychology and economics: Evidence from the field', *Journal of Economic Literature* **47**(2), 315–372.

Fasolo, B., Carmeci, F. A. & Misuraca, R. (2009), 'The effect of choice complexity on perception of time spent choosing: When choice takes longer but feels shorter', *Psychology & Marketing* **26**(3), 213–228.

Fisch, J. E. & Wilkinson-Ryan, T. (2014), 'Why do retail investors make costly mis- takes? an experiment on mutual fund choice', *University of Pennsylvania Law Review* **162**(3), 605–647.

Freedman, D. A. (2008), 'On regression adjustments to experimental data', *Advances in Applied Mathematics* **40**(2), 180–193.

Freedman, D. A. et al. (2008), 'On regression adjustments in experiments with several treatments', *The annals of applied statistics* **2**(1), 176–196.

Friesen, L. & Earl, P. E. (2015), 'Multipart tariffs and bounded rationality: an experimental analysis of mobile phone plan choices', *Journal of Economic Behavior & Organization* **116**, 239–253.

Golman, R., Hagmann, D. & Loewenstein, G. (2017), 'Information avoidance', *Journal of Economic Literature* **55**(1), 96–135.

Gourville, J. T. & Soman, D. (2005), 'Overchoice and assortment type: When and why variety backfires', *Marketing science* **24**(3), 382–395.

Greifeneder, R., Scheibehenne, B. & Kleber, N. (2010), 'Less may be more when choosing is difficult: Choice complexity and too much choice', *Acta psychologica* **133**(1), 45–50.

Harstad, R. M. & Selten, R. (2013), 'Bounded-rationality models: tasks to become intellectually competitive', *Journal of Economic Literature* **51**(2), 496–511.

Hartzmark, S. M. & Sussman, A. B. (2019), 'Do investors value sustainability? a natural experiment examining ranking and fund flows', *The Journal of Finance* **74**(6), 2789–2837.

Haugtvedt, C. P., Herr, P. M. & Kardes, F. R. (2018), Consumer decision making: A choice goals approach: James r. bettman, mary frances luce, and john w. payne, *in* 'Handbook of Consumer Psychology', Routledge, pp. 588–609.

Hensher, D. A. (2010), 'Attribute processing, heuristics and preference construction in choice analysis', *State-of Art and State-of Practice in Choice Modelling, Emerald Press, UK* pp. 35–70.

Hortaçsu, A. & Syverson, C. (2004), 'Product differentiation, search costs, and competition in the mutual fund industry: A case study of s&p 500 index funds', *The Quarterly journal of economics* **119**(2), 403–456.

Hsee, C. K. & Zhang, J. (2010), 'General evaluability theory', *Perspectives on Psychological Science* **5**(4), 343–355.

Imbens, G. W. & Rubin, D. B. (2015), *Causal inference in statistics, social, and biomedical sciences*, Cambridge University Press.

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.

Jang, J. M. & Yoon, S. O. (2016), 'The effect of attribute-based and alternative-based processing on consumer choice in context', *Marketing Letters* **27**(3), 511–524.

Jia, J., Luce, M. F. & Fischer, G. W. (2004), *Consumer Preference Uncertainity: Measures of Attribute Conflict and Extremity*, Office of Research, Singapore Management University.

Kalaycı, K. & Serra-Garcia, M. (2012), 'Complexity and biases: An experimental study'.

Kempf, A. & Ruenzi, S. (2006), 'Status quo bias and the number of alternatives: An empirical illustration from the mutual fund industry', *The journal of behavioral finance* **7**(4), 204–213.

Lin, W. (2013), 'Agnostic notes on regression adjustments to experimental data: Reexam- ining freedmans critique', *The Annals of Applied Statistics* **7**(1), 295–318.

Loewenstein, G. & Lerner, J. S. (2003), 'The role of affect in decision making.'.

Lucas Jr, R. E. (1986), 'Adaptive behavior and economic theory', *journal of Business* pp. S401–S426.

Mcfadden, D. (1997), 'Rationality for economists presented at the nsf symposium on eliciting preferences', *Berkeley, California* .

Mourali, M. & Pons, F. (2009), 'Regulatory fit from attribute-based versus alternative-based processing in decision making', *Journal of Consumer Psychology* **19**(4), 643–651.

Oprea, R. (2020), 'What makes a rule complex?', *American economic review* **110**(12), 3913–51.

Phatak, N. (2012), 'Menu-based complexity: Experiments on choice over lotteries'.

Phillips, N. D. (2017), 'Yarrr! the pirates guide to r', *APS Observer* **30**(3).

Rabin, M. (2013), 'Incorporating limited rationality into economics', *Journal of Economic Literature* **51**(2), 528–43.

Rees-Jones, A. & Taubinsky, D. (2016), *Heuristic perceptions of the income tax: Evidence and implications for debiasing*, National Bureau of Economic Research.

Scheibehenne, B., Greifeneder, R. & Todd, P. M. (2010), 'Can there ever be too many options? a meta-analytic review of choice overload', *Journal of consumer research* **37**(3), 409–425.

Scholl, B. & Fontes, A. (2022), 'Mutual fund knowledge assessment for policy and decision problems', *Financial Services Review (forthcoming)* .

Scholl, B., Silverman, D. & Enriquez, M. (2020*a*), 'Disclosure complexity and fund performance', *Working Paper, SEC Office of the Investor Advocate* .

Scholl, B., Silverman, D. & Enriquez, M. (2020*b*), 'Text complexity and performance', *Rand BeFi Presentation* .

Scholl, B. & VanEpps, E. (2022), 'The you get what you pay for heuristic'.

Sethi-Iyengar, S., Huberman, G. & Jiang, W. (2004), 'How much choice is too much? contributions to 401 (k) retirement plans', *Pension design and structure: New lessons from behavioral finance* **83**, 84–87.

Thaler, R. H., Sunstein, C. R. & Balz, J. P. (2013), 'Choice architecture', *The behavioral foundations of public policy* pp. 428–439.

Tse, A., Friesen, L. & Kalaycı, K. (2016), 'Complexity and asset legitimacy in retirement investment', *Journal of Behavioral and Experimental Economics* **60**, 35–48.

Tsiatis, A. A., Davidian, M., Zhang, M. & Lu, X. (2008), 'Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach', *Statistics in medicine* **27**(23), 4658–4677.

Tversky, A. & Shafir, E. (1992), 'Choice under conflict: The dynamics of deferred decision', *Psychological science* **3**(6), 358–361.

Venkatraman, V., Payne, J. W., Bettman, J. R., Luce, M. F. & Huettel, S. A. (2009), 'Separate neural mechanisms underlie choices and strategic preferences in risky decision making', *Neuron* **62**(4), 593–602.

Widing, R. E. & Talarzyk, W. W. (1993), 'Electronic information systems for consumers: an evaluation of computer-assisted formats in multiple decision environments', *Journal of Marketing Research* **30**(2), 125–141.