REVEALING CHOICE BRACKETING

ANDREW ELLIS AND DAVID J. FREEMAN

Abstract. Experiments suggest that people fail to take into account interdependencies between their choices – they do not broadly bracket. Researchers often instead assume that people narrowly bracket, but existing designs do not test it. We design a novel experiment and revealed preference tests for how someone brackets their choices. In portfolio allocation under risk, social allocation, and induced-value shopping experiments, 40-43% of subjects are consistent with narrow bracketing and 0-16% with broad bracketing. Adjusting for each model's predictive precision, 74% of subjects are best described by narrow bracketing, 13% by broad bracketing, and 6% by intermediate cases.

JEL codes: D01, D90

Keywords: choice bracketing, individual decision-making, revealed preference, experiment.

Individuals face many interconnected decisions. How an individual takes into account the interdependencies when choosing, or how they *bracket* these choices, significantly influences their decision-making process. Bracketing determines which outcomes are evaluated as gains or losses and fair or unfair, and also plays a role in measuring parameters like risk aversion.

Date: March 2024.

Freeman thanks UCSB and QMUL for hosting him while parts of this paper were written, and UToronto for use of its TEEL lab for our experiments. We especially thank Johannes Hoelzemann for help in coordinating experiments at TEEL. We thank Pietro Ortoleva, Stefano DellaVigna, three anonymous referees, Ted Bergstrom, Ignacio Esponda, Erik Eyster, Daniel Gottlieb, Yoram Halevy, Alex Imas, Zihan Jia, Matt Levy, Marco Mariotti, Antony Millner, Ryan Oprea, Luba Petersen, Kate Smith, Balazs Szentes, Emanuel Vespa, Sevgi Yuksel, Lanny Zrill, and audiences at UCSB, Berlin, Chapman, QMUL, UCL, UTS, Berkeley/Booth, UCSD, Ottawa, UC Davis, D-TEA 2020, ESA 2020, RUD 2020, ESNAWM 2021, MBEES 2021, CEA 2021, BRIC 2022 for helpful discussions and suggestions. We thank Alex Ballyk, Daphne Baldassari, Matheus Thompson Bandeira, Priscilla Fisher, En Hua Hu, Louise Song, Johnathan Wang, and Marie Zagre for research assistance. This paper was funded by a 2018 SFU/SSHRC Institutional Grant and SSHRC Insight Grant 435-2019-0658 and was conducted under SFU Research Ethics Study #2016s0380.

Ellis: Department of Economics, London School of Economics. e-mail: a.ellis@lse.ac.uk. Web: `http://personal.lse.ac.uk/ellisa1/`.

Freeman: Department of Economics, Simon Fraser University. e-mail: david_freeman@sfu.ca. Web: `http://www.sfu.ca/~dfa19/`.

Nearly every behavioral model and most "rational" ones require some assumption about how people bracket choices.

There are many ways to bracket. Optimal decision-making requires that people *broadly bracket*, considering every feasible combination of choices and selecting the best. The most common alternative is that people *narrowly bracket* by making each decision without considering any interdependencies. However, these two extremes are far from exhaustive. For instance, Barberis *et al.* (2006) and Rabin & Weizsäcker (2009) propose a hybrid of the two called *partial-narrow bracketing*.

Most experimental evidence interpreted as being *for* narrow bracketing is actually evidence *against* broad bracketing. This evidence, surveyed in Section 2, comes mainly from studies that follow a similar design to Tversky & Kahneman (1981, Problem 3) or Kahneman & Tversky (1979, Problems 11-12). In the former, each subject makes two concurrent choices, and one pair of choices generates a distribution over outcomes dominated by another pair. They find that many subjects choose the dominated pair. In the latter, two groups of subjects face choices between lotteries that are economically identical but differ in how payments are divided between an endowed income and an active choice. The two groups make different choices. Both designs provide evidence against broad bracketing. However, narrow bracketing makes no testable predictions in either design: any choices are consistent with it. This leaves open the question of whether narrow bracketing is a good description of behavior.

We propose a theoretical framework and experimental design to test how an individual brackets. Every subject makes several decisions, each comprised of one or more parts. Their payoff from a decision is determined by the sum of the items chosen in its parts. A subject is consistent with narrow bracketing if they maximize a preference relation in each part, and consistent with broad bracketing if they optimize after integrating all parts of the decision into a single feasible set. We characterize all the testable predictions for narrow, broad, and partial-narrow bracketing. For example, we check for narrow bracketing by performing a standard revealed preference test on a dataset of their choices that treats each part as an independent observation. The results provide individual-level, non-parametric tests that rely only on monotonicity of the underlying preferences. They provide the theoretical basis for an experimental design that tests all three models of choice bracketing.

To implement these tests, we conduct experiments in which subjects make five decisions consisting of one or two parts. Every part is a budget set, and integrating the choices only requires addition. Unlike the existing designs described above, narrow and broad bracketing make distinct predictions in each of the two-part decisions. Every good in the first part is a perfect substitute for a good in the second. Broad bracketers recognize and take full advantage of this substitutability, while narrow bracketers do not.

We apply our design and tests to portfolio choices among risky assets (Choi *et al.*, 2007), dividing money between two anonymous other subjects (Andreoni & Miller, 2002; Fisman *et al.*, 2007), and standard consumer problems with induced values over bundles. The last allows us to conduct even more powerful tests of bracketing because preferences are known. Across the three experiments, 40-43% of subjects are consistent with narrow bracketing, 0-16% are consistent with broad, 1-32% are consistent with partial-narrow but neither broad nor narrow, and 24%-44% are consistent with none of the three. No subjects are consistent with *both* broad and narrow bracketing. We then determine which model best describes each subject's bracketing based on both the number of errors they make relative to each model and the predictive precision of that model (Selten, 1991). We classify 6-13% of subjects to broad bracketing, 68-78% to narrow bracketing, and only 3-9% to partial narrow bracketing, with variation across experiments.[1]

Failing to bracket broadly has consequences for welfare. Economists as early as Smith (1776) argue that agents benefit from specialization. A broad bracketer specializes by purchasing as much of a good as possible when it is relatively cheap before purchasing any when it has a higher opportunity cost. This increases utility in the same way that comparative advantage increases total productivity in classic trade models.[2] Broad bracketers apply Smith's insights about the productivity increases from specialization to individual decision-making, while narrow bracketers fail to reap the gains from specialization.[3] Our approach

[1]The remaining subjects are poorly described by all models and left unclassified.

[2]Baron & Kemp (2004) provide a survey measure of understanding of specialization, and show it is correlated with attitudes to free trade.

[3]We thank a referee for pointing to this analogy.

allows us to quantify the magnitude of these losses. For instance, the subjects classified as narrow bracketers made an average of $1.29 less than the broad bracketers from the two-part decisions in the Shopping Experiment, more than 10% of the variable payment.

Because our tests are individual-level, they can provide evidence on why so many people bracket narrowly despite the gains from broad bracketing. In online follow-up experiments, we recorded choice process data and implemented a nudge to encourage subjects to "examine both parts" of the decision. While the nudge increased the proportion of subjects who looked at both parts, it had a limited effect on the rate of broad bracketing. In one arm of the Online Risk Experiment, we observed that about a quarter of the narrow bracketers had enough information to bracket broadly but did not do so. In the Online Shopping Experiment, a similar fraction used a calculator to compute the payoff of one or more bundles that were only feasible at the decision level and not in any part on its own. Both these behaviors indicate some consideration of how choices from both parts combine to determine payoffs. This suggests that not all narrow bracketing results from simply looking at one part of a decision at a time and making an optimal choice for that part on its own. Instead, it suggests that some individuals consider both decision problems, yet ultimately implement choices that are only compatible with narrow bracketing.

Bracketing also affects how one interprets behavior. A narrow bracketer's reluctance to take a small yet actuarially favorable gamble indicates only slight risk aversion. However, for broad bracketers, the rejection of the same gamble signifies extreme risk aversion, given its minimal impact on their overall wealth (Rabin, 2000). Our approach enables us to separate underlying preferences from bracketing, with implications for measuring risk or inequity aversion.

This separation allows us to measure heterogeneity in bracketing directly and to account for it when inferring preferences. To illustrate the importance of doing so, consider the Social Experiment. In Part 2 of Decision 1 (see Table 2), subjects on average allocate unequally between the two anonymous others, dividing $16 into $6.80 and $9.20. If all subjects bracketed narrowly, this would suggest a lack of concern for equity. However, we classify 10% of subjects as broad bracketers and 75% as narrow bracketers. The narrow bracketers on average allocate $7.75 and $8.25 to each person, and the broad bracketers all achieve perfectly equal allocations in the decision by allocating a $14 - $2 split in that part and a $0 - $12 split in the other. After taking bracketing into account, all these subjects are consistent with inequity aversion. The only existing work that estimates heterogeneity in bracketing, Rabin & Weizsäcker (2009), estimates a between-subjects, parametric, structural model and assumes that all subjects have the same preferences. In contrast, our approach is individual-level, non-parametric, and allows heterogeneity in preferences.

Most applications of prospect theory explicitly assume narrow bracketing (Camerer 2004, Table 5.1; O'Donoghue & Sprenger, 2018, Sections 3.4 and 4.5). Since it is impossible to avoid all other risks, a broad bracketer will not exhibit noticeable small-stakes loss aversion (Barberis *et al.*, 2006). At the same time, other applications of prospect theory, such as casino gambling (Barberis, 2012; Ebert & Strack, 2015), require broader bracketing. Our results suggesting heterogeneity in bracketing may reconcile the results rejecting broad bracketing (e.g. Tversky & Kahneman, 1981) and those more supportive of it (e.g. Heimer *et al.*, 2020 and Baillon *et al.*, 2022).

Choice bracketing is also important for social decisions. For example, Sobel (2005, p. 400) remarks that in experiments studying social decisions, a subject should "maximize her monetary payoff in the laboratory and then redistribute her earnings to deserving people later... if the concern for inequity was 'broadly bracketed.'" However, the literature he surveys and our analysis both find evidence consistent with subjects caring about narrowly-bracketed equity (e.g. Charness & Rabin, 2002; Bolton & Ockenfels, 2000).

Standard economic analyses make assumptions about bracketing as well. As noted by a recent behavioral economics textbook, "Although choice bracketing has been largely ignored in economics, models in economic theory often implicitly assume and invoke it" (Dhami, 2016, p. 1469). Many analyses exclude some related choices from the model yet interpret results in terms of a fully rational choice. Our results suggesting that most people are narrow bracketers lend support to this approach, with the caveat that some seem best modeled as broad bracketers.

1. Testing choice bracketing

We consider a decision-maker (DM) who faces $T \geq 1$ decision problems involving alternatives contained in $\mathbb{R}^n_+$. Decision t consists of $K_t \geq 1$ parts. Formally, the feasible set of part k from decision t is $B^{t,k}$, assumed to be compact and non-empty. The DM chooses one alternative, denoted $x^{t,k}$, from each part, $B^{t,k}$. Thus, the analyst observes the *dataset*

$$\mathcal{D} = \left\{ \left(x^{t,k}, B^{t,k} \right) \right\}_{(t,k)}$$

indexed by decisions and parts and where $x^{t,k} \in B^{t,k}$. They face parts concurrently, and their payoff in decision t depends only on the sum of their choices across parts of the decision,

$$x^t = \sum_{k=1}^{K_t} x^{t,k},$$

which we call the "final alternative." A real world analogue consists of a scanner dataset with purchases at different stores (parts, indexed by k) in a given time period (a decision, indexed by t), and the shopper consumes these purchases after buying at all stores but before the next period.

A DM *broadly brackets* if they maximize an increasing utility function over the set of feasible final alternatives for each decision problem,

$$B^t = \left\{ \sum_{k=1}^{K_t} y^{t,k} : y^{t,k} \in B^{t,k} \right\}.$$

In contrast, they *narrowly bracket* if they choose the alternative in part k of decision t that maximizes an increasing utility function over $B^{t,k}$.[4] Both require that the DM is "rational" in that they maximize some well-behaved preference relation, but neither requires any further assumptions about preferences beyond monotonicity. A narrow bracketer optimizes part-by-part, while a broad bracketer does so decision-by-decision.

Our goal is to test which, if any, models of bracketing can explain a subject's choices. Formally, a dataset $\mathcal{D}$ is *rationalized by broad bracketing* if there exists an increasing utility

[4] A narrow bracketer acts as if they perceive alternatives correctly and maximizes a well-behaved preference over them, but misperceives the budget set. In contrast, a DM who misperceived correlation, e.g. Eyster & Weizsäcker (2016) or Ellis & Piccione (2017), perceives the choice set correctly but misperceives the alternatives themselves.

function $u : \mathbb{R}^n_+ \to \mathbb{R}$ so that

$$x^t = \arg\max_{x \in B^t} u(x)$$

for every decision t, and a dataset $\mathcal{D}$ is *rationalized by narrow bracketing* if there exists an increasing $u : \mathbb{R}^n_+ \to \mathbb{R}$ so that

$$x^{t,k} = \arg\max_{x \in B^{t,k}} u(x)$$

for every part k and decision t.[5] The next two subsections provide necessary and sufficient conditions for $\mathcal{D}$ to be rationalized by either of the two.

1.1. **Predictions.** We first provide necessary conditions for rationalization by broad and narrow bracketing. The first two predictions build on the observation that the DM's choice reveals their preference among the set of alternatives over which they optimize. The predictions combine the Weak Axiom of Revealed Preference (WARP) with a form of bracketing.

Prediction 1 (NB-WARP)**.** Suppose that $\mathcal{D}$ is rationalized by narrow bracketing.
If $\left(x^{t,k}, B^{t,k}\right), \left(x^{t',k'}, B^{t',k'}\right) \in \mathcal{D}$ and $x^{t',k'} \in B^{t,k} \subseteq B^{t',k'}$, then $x^{t,k} = x^{t',k'}$.

Prediction 2 (BB-WARP)**.** Suppose that $\mathcal{D}$ is rationalized by broad bracketing.
For decisions t, t' in $\mathcal{D}$, if $x^{t'} \in B^t \subseteq B^{t'}$, then $x^t = x^{t'}$.

Narrow bracketing requires that WARP holds when comparing any pair of parts of decisions, even when they belong to economically different decisions. Broad bracketing implies that WARP holds at the decision level, comparing final alternatives that are feasible in both aggregate budget sets.

The next prediction reflects the appropriate manifestation of monotonicity in our setting.

Prediction 3 (BB-Mon)**.** Suppose that $\mathcal{D}$ is rationalized by broad bracketing.
For any decision t in $\mathcal{D}$ and any $y \in B^t$, if $y \geq x^t$, then $y = x^t$.

BB-Mon requires that the subject chooses on the frontier of their aggregate budget set in a given decision.[6]

[5]These definitions assume that each choice represents strict preference, but one can easily generalize them and the results that follow to allow for indifference.
[6]Narrow bracketing predicts a similar condition at the part level. Our experimental implementation forces it to hold, so we do not formally include it here.

1.2. **Characterization.** The above predictions are necessary but not sufficient conditions for each type of bracketing. We obtain tight characterizations by extending the logic of the above to include indirect implications, in the same manner that Strong Axiom of Revealed Preference (SARP) extends WARP. Theorem 1 shows that these are necessary and sufficient conditions for a given dataset to be rationalized by a particular form of bracketing.

Our tests are based on applying SARP to an ancillary dataset. Say that a bundle x is *directly revealed preferred to* y in a dataset $\mathcal{D}'$ consisting of single-part decisions, written $xP^{\mathcal{D}'}y$, if either $x \geq y$ and $x \neq y$ or there exists $(x, B) \in \mathcal{D}'$ so that $y \in B\backslash\{x\}$. A dataset $\mathcal{D}'$ *satisfies SARP* if the binary relation $P^{\mathcal{D}'}$ is acyclic.

We say that say that $\mathcal{D}$ satisfies BB-SARP if the ancillary dataset

$$\mathcal{D}^{BB} = \left\{\left(x^t, B^t\right)\right\}_{t \in T}.$$

satisfies SARP. Similarly, we say that say that $\mathcal{D}$ satisfies NB-SARP if SARP is satisfied by the ancillary dataset

$$\mathcal{D}^{NB} = \{(x^s, B^s)\}_{s=1}^{\sum_{t=1}^{T} K_t}$$

where for each part (t, k), there exists a unique s so that $(x^s, B^s) = \left(x^{t,k}, B^{t,k}\right)$. In $\mathcal{D}^{NB}$, each part of every decision is treated as a separate, independent observation, whereas each decision is in $\mathcal{D}^{BB}$.

Theorem 1. *The following are true:*

(i) The dataset $\mathcal{D}$ satisfies BB-SARP if and only if $\mathcal{D}$ is rationalizable by broad bracketing, and

(ii) The dataset $\mathcal{D}$ satisfies NB-SARP if and only if $\mathcal{D}$ is rationalizable by narrow bracketing.

Theorem 1 characterizes the complete testable implications of broad and narrow bracketing. Moreover, both conditions are readily applied using standard computational tools.[7] One can make tighter predictions by imposing more structure on the utility function that

[7]We note that B^t typically has a piece-wise linear budget-line even if each $B^{t,k}$ is a standard budget set.

rationalizes the data. For instance, we impose that preferences are symmetric when we apply these tests to our experiments. This enables stronger versions of the two tests.[8]

1.3. **Partial-narrow bracketing.** We also consider an intermediate model of bracketing lying between the two extremes, called *α-partial-narrow bracketing (PNB).* Inspired by Barberis *et al.* (2006), Barberis & Huang (2009), and Rabin & Weizsäcker (2009), the dataset $\mathcal{D}$ is *rationalized by α-partial-narrow bracketing* if

$$\left(x^{t,1},\ldots,x^{t,K_t}\right)=\arg\max_{y^{t,k}\in B^{t,k}\ \forall k}\alpha\sum_{k=1}^{K^t}u\left(y^{t,k}\right)+(1-\alpha)\,u\left(\sum_{k=1}^{K^t}y^{t,k}\right) \tag{1}$$

for all t.[9] A PNB DM's choices maximize a weighted average of the utility of the final alternative and of the utility of their choice in each part.

A broad bracketer takes into account that good i in part k is a perfect substitute for good i in part k', whereas a narrow bracketer does not consider any complementarities across parts of the decision.[10] In contrast, a PNB DM views the same good in different parts as imperfect substitutes for each other. They act as if they make a single choice in decision t of an element of $\mathbb{R}_+^{nK_t}$, but they maximize a utility function derived from u rather than u itself. In contrast, a broad bracketer acts as if they choose a single element of $\mathbb{R}_+^n$, and a narrow bracketer as if they make K_t independent choices of elements of $\mathbb{R}_+^n$.[11]

[8]Specifically in $\mathbb{R}_+^2$, if $(x,y)\,P^{\mathcal{D}}\,(x',y')$, then $(y,x)\,P^{\mathcal{D}}\,(x',y')$, $(x,y)\,P^{\mathcal{D}}\,(y',x')$, and $(y,x)\,P^{\mathcal{D}}\,(y',x')$. In Appendix B, we derive some immediate implications of symmetry.

[9]The first two papers consider an average of the certainty equivalents, $[u^{-1}(E[u(x^t)]) + b_0\sum_{k'=1}^{K^t}v^{-1}(E[v(x^{t,k})]).]$ Unlike our specification, they allow for the narrow utility function to differ from the broad utility function and assume expected utility. The algorithm in Theorem 2 can be adapted to allow for different narrow and broad utility functions ($u\neq v$) at the cost of less predictive power. Our specification is closest to Rabin & Weizsäcker (2009, p. 1513), although they also assume expected utility.

[10]Kőszegi & Matějka (2020) also base their approach to mental budgeting on substitutability between parts.

[11]Recent models by Vorjohann (2020) and Zhang (2023) have this feature as well, but with different utilities over $\mathbb{R}_+^{nK_t}$ than Equation (1). They capture departures from broad bracketing by either substituting in a reference point for other parts or by imposing separability where u is inseparable. In a two part decisions, these models are as follows. Vorjohann (2020) considers a DM who maximizes $v\left(x^{t,1},x^{t,2}\right)=u\left(x^{t,1},r^2\right)+u\left(r^1,x^{t,2}\right)$ for reference points $r^1\in B^{t,1}$ and $r^2\in B^{t,2}$, instead of $v^{BB}\left(x^{t,1},x^{t,2}\right)=u\left(x^{t,1},x^{t,2}\right)$. Zhang (2023), for the specification most similar to our setting, considers a DM who maximizes $U\left(x^{t,1},x^{t,2}\right)=E_{x^{t,1}}\left[u\left(z+u^{-1}\left(E_{x^{t,2}}\left[u(y)\right]\right)\right)\right]$ instead of $U^{BB}\left(x^{t,1},x^{t,2}\right)=E_{x^{t,1}+x^{t,2}}\left[u\left(z\right)\right]$, where E_x is the expectation operator with respect to the lottery corresponding to x.

An algorithm *tests* whether $\mathcal{D}$ is rationalizable by α-partial-narrow bracketing if it inputs a dataset $\mathcal{D}$ and (deterministically) outputs 1 when $\mathcal{D}$ is rationalizable by α-partial-narrow bracketing and outputs 0 otherwise.

Theorem 2. *There exists an algorithm that tests whether $\mathcal{D}$ is rationalizable by α-partial-narrow bracketing.*

We construct the algorithm in Appendix A. The algorithm reduces the problem of determining whether $\mathcal{D}$ is consistent with α-partial-narrow bracketing to that of a standard linear programming problem. It runs in polynomial time with a fixed set of possible outcomes of each decision and its parts. We elaborate on the ideas behind it, and provide a formal proof in the Appendix.

The algorithm takes advantage of the formal similarity between Equation (1) and expected utility. Specifically, an $\alpha-$PNB DM with utility u chooses $x^{t,1}$ in part 1 and $x^{t,2}$ in part 2 if and only if an expected utility DM with Bernoulli index u chooses $(\psi\alpha, x^{t,1}; \psi\alpha, x^{t,2}; \psi(1-\alpha), x^t)$ from the menu of lotteries $\{(\psi\alpha, y^1; \psi\alpha, y^2; \psi(1-\alpha), y^1+y^2) : y^1 \in B^{t,1} \ \& \ y^2 \in B^{t,2}\}$ where $\psi = (1+\alpha)^{-1}$. The algorithm first transforms the original dataset to an ancillary dataset where each decision is a single choice from a menu of lotteries. Each lottery in the menu is structured so that an alternative in part k occurs with probability proportional to α for each part k and their sum occurs with probability proportional to $(1-\alpha)$. The DM's choice is the lottery over their choices in each part and their sum. Existing results, e.g. (Clark, 1993), provide necessary and sufficient conditions for the ancillary dataset to be rationalizable by expected utility. We show that whether or not it can be rationalized can be determined by examining the solution to a linear program.

We also consider a related intermediate model of bracketing, "PNB-Personal Equilibrium (PNB-PE)." Formally,

$$x^{t,k} = \underset{y^{t,k}\in B^{t,k}}{\arg\max}\left[\alpha u\left(y^{t,k}\right) + (1-\alpha)u\left(\sum_{k'\in\{1\le k'\le K^t:k'\neq k\}} x^{t,k'} + y^{t,k}\right)\right].$$

for every part k of decision t. While the functional form is similar to PNB, the DM acts as if they make K_t independent choices of elements of $\mathbb{R}^n_+$, as a narrow bracketer would would.

Their choices are the personal equilibrium (PE) of an intrapersonal game (Kőszegi & Rabin, 2006; Lian, 2019). A PNB-PE DM only considers the utility effect of $x^{t,-k}$ when choosing $x^{t,k}$, ignoring that they could change $x^{t,-k}$ in response to their choice. A similar algorithm to the one that tests α-PNB can test α-PNB-PE.

2. Existing Evidence on Bracketing

The models we consider capture different forms of what Read *et al.* (1999) call choice bracketing: how a person groups "individual choices together into sets." We first discuss three strands of research into static, concurrent decisions to which our setting directly applies before turning to the important case of dynamic, or temporal, bracketing.

Most existing direct evidence of non-broad choice bracketing is derived from Tversky & Kahneman (1981), in which subjects face the decision problem described in Table 1. This

Table 1. Tversky and Kahneman's Decision problem

Imagine that you face the following pair of concurrent decisions. First, examine both decisions, then indicate the options you prefer.

Decision (i). Choose between:	
A. a sure gain of \$240 [85%]	B. 25% chance to gain \$1000, and 75% chance to gain nothing [16%]
Decision (ii). Choose between:	
C. a sure loss of \$750 [13%]	D. 75% chance to lose \$1000, and 25% chance to lose nothing [87%]

design fits into our theoretical setting as a single-decision dataset with $B^{1,1} = \{A, B\}$ and $B^{1,2} = \{C, D\}$. The choice combination (A, D) made by 73% of their subjects generates a first-order stochastically-dominated distribution over outcomes compared to the pair of choices (B, C) and so violates BB-Mon and BB-SARP. Their design cannot falsify narrow bracketing since every combination of choices satisfies NB-SARP. Without further restrictions, their results are uninformative about the choice bracketing of subjects who do not choose A and D. Yet about 70% of subjects made non (A, D) choices in incentivized follow-up experiments (Rabin & Weizsäcker, 2009; Koch & Nafziger, 2019).

A related set of experiments considers how an exogenously fixed quantity, like a monetary endowment, an asset, or a background risk, affects a person's choice when separated from the description of available alternatives.[12] For instance, studies suggest subjects behave as-if they do not incorporate their endowment in risk-taking choices (Kahneman & Tversky, 1979, Problems 11-12), in social allocation tasks (Exley & Kessler, 2018), and in labor supply choices (Fallucchi & Kaufmann, 2021). These designs fit into our theoretical framework as decisions in which one part is a singleton set (the endowment) and one part a non-singleton set (the active choice), and these papers provide evidence against broad bracketing.

The final body of evidence shows how failures of fungibility affect choice (Thaler, 1999). Studies of the "flypaper effect" suggest that transfers earmarked for a particular type of spending tend to actually be spent there (Hines & Thaler, 1995; Abeler & Marklein, 2017). Empirical evidence from consumption choices after unexpected price changes support the lack of fungibility (Hastings & Shapiro, 2013, 2018).[13]

This evidence has a more complicated relationship with our setting. Thaler (1985; 1999) and others (Galperti, 2019; Kőszegi & Matějka, 2020) explain the evidence through mental accounting (or budgeting). At its most general, this refers to a person breaking up the overall decision into categories. In contrast, narrow choice bracketing is a failure to aggregate smaller parts into a larger decision. These are not mutually exclusive. If categories and parts coincide, our model of narrow choice bracketing provides an extreme model of mental accounting. But in general, mental accounting is consistent with either broadly-bracketing across parts (Thaler, 1985, p 207) or with narrowly bracketing each part (e.g. as described by Corollary 1 of Kőszegi & Matějka, 2020).[14]

Choice bracketing is also relevant in dynamic settings. Gneezy & Potters (1997) show that subjects make different choices period-by-period than when they must choose in advance, evidence of non-broad bracketing.[15] In contrast, recent evidence from Heimer *et al.*

[12]Related theoretical results by Samuelson (1963), Rabin (2000), Safra & Segal (2008), and Mu *et al.* (2020) suggest that behavior in small-stakes bets cannot be reconciled with behavior in large-stakes bets and broad bracketing. The results of Gneezy & Potters (1997); Benartzi & Thaler (1995) provide support for their theoretical predictions.

[13]Evers *et al.* (2022) suggest that similarity mediates how people group outcomes into categories.

[14]One can extend our framework by replacing the "SARP" part of NB- or BB- SARP with a revealed preference test for mental accounting, such as that of Blow & Crawford (2018). Then, we have two joint tests: one for mental accounting where the budgets for each category are determined part-by-part, and one for mental accounting where the budgets for each category are determined at the decision-level.

[15]The comparison in Cox *et al.* (2015) of their "One Task" treatment and their "Pay all sequentially" treatment provides between-subjects evidence that fails to reject the null that every subject narrowly brackets.

(2020) (building on Barberis (2012); Ebert & Strack (2015)) suggests that subjects behave as if they do not narrowly bracket future risk-taking opportunities. In financial decision-making, Shefrin & Statman (1985) and Odean (1998) document the disposition effect, and Thaler & Johnson (1990) and Imas (2016) document the house money effect. Either effect is inconsistent with both fully narrow and fully broad bracketing.

Our approach can be generalized to a dynamic setting, but two key complications arise. First, the setting should comprise decision trees rather than sets of alternatives. Second, behavior may be dynamically inconsistent. Our approach can be adapted but would apply only under dynamic consistency. Future work should incorporate dynamic inconsistency.

We note that there are many ways to fail to bracket broadly in a dynamic setting. For instance, an agent may only look backwards and take into account the results of their past decisions but neglect to consider how their current decision interacts with their future ones. Or they may only look forwards and take into account future interactions but not past results. Even more combinations are possible: they could look neither forwards nor backwards, or only look forwards a certain number of periods.

3. Experimental Design

We design and conduct three experiments to test the models of bracketing in different domains of choice. In each experiment, a participant faces five decision rounds, each consisting of one or two parts.[16] Each part consists of all feasible integer-valued bundles of two goods obtained from a linear (or in one case, a piece-wise linear) budget set. At the end of the experiment, exactly one round is randomly selected for payment, which we call the "round that counts". We sum all goods purchased in all parts of the decision in the round that counts to obtain the final bundle that determines payments.[17] By design, there are no complementarities across decisions. We implement this experimental design to study choice

They note that this is consistent with prior literature that finds little evidence for "wealth effects" (i.e. violations of narrow bracketing) in economics experiments that pay all choices sequentially.

[16]Compared to choice-from-budget-set experiments like Choi *et al.* (2007), each subject faces fewer rounds of decisions in our experiment and each budget set is coarser to allow us to conduct the experiment on paper. We made this design choice to minimize potential for learning to bracket, and also to slow down subjects' progression through the experiment to encourage slow and deliberate decision-making.

[17]In the Social Experiment we modify this procedure slightly: one person-decision pair per anonymous group of two subjects is randomly selected to determine payments of another anonymous group of two subjects.

bracketing in three domains of interest: portfolio choice under risk (Risk), a social allocation task (Social), and a consumer choice experiment in which we induced subjects' values (Shopping).

In the Risk Experiment, each part of every decision asks the subject to choose an integer allocation of tokens between two assets. Each asset pays off on only one of two equally likely states: Asset A (or C) pays out only on a die roll of 1-3 whereas Asset B (or D) pays out only on a die roll of 4-6. The payoff of each asset varies across decision problems and across parts. Because each decision problem uses assets with two equally likely states, preferences over portfolios of monetary payoffs for each state should be symmetric across states.

In the Social Experiment, each part of every decision asks the subject to choose an integer allocation of tokens between two anonymous other subjects, Person A and Person B. The value of each token to A and B varies across decision problems and across parts. Because the two recipients are anonymous, we expect preferences to be symmetric across money allocated to A versus B.

In the Shopping Experiment, each part of every decision asks the subject to choose a bundle of integer quantities of fictitious "apples" and "oranges" subject to a budget constraint. The monetary payment for the experiment is calculated from the final bundle in the round that counts according to the function $\$\text{pay} = \frac{2}{5}\left(\sqrt{\#\text{apples}} + \sqrt{\#\text{oranges}}\right)^2$.[18] This function induces payoffs that are symmetric in apples and oranges and that are strictly variety-seeking. Any subject who prefers more money to less will wish to maximize this payoff function regardless of their underlying utility function.

Our experimental budget sets, summarized in Table 2, allow us to conduct our revealed preference tests in the Risk and Social Experiments, and to conduct analogous tests that make use of the induced value function in our Shopping Experiment. Throughout, we refer to part k of decision t as Dt.k, or Dt if a round has only one part. The order of decisions and parts was varied across sessions (Table 15 in Online Supplement H).

[18]Subjects were provided with a payoff table (Online Supplement H, Figure 18) to calculate the earnings that would result from any possible final bundle, so could maximize earnings without having to manually compute this function.

		Risk			Social			Shopping		
Decision	Part	I	Asset A/C	Asset B/D	I	V_A	V_B	I	p_a	p_o
D1	1	10	(\$1, \$0)	(\$0, \$1.20)	10	\$1	\$1.20	8	2	1
	2	16	(\$1, \$0)	(\$0, \$1)	16	\$1	\$1	24	2	2
D2	1	14	(\$2, \$0)	(\$0, \$2)	14	\$2	\$2	32	2	1 (for 1st 8), 2
D3	1	10	(\$1, \$0)	(\$0, \$1)	10	\$1	\$1	30	3	3
	2	10	(\$1, \$0)	(\$0, \$1.20)	10	\$1	\$1.20	24	3	2
D4	1	16	(\$1, \$0)	(\$0, \$1)	16	\$1	\$1	12	1	1
D5	1	10	(\$1, \$0)	(\$0, \$1.20)	10	\$1	\$1.20	48	6	4

I: income for a part (in tokens)

(\$$x$, \$$y$) indicates one unit of asset pays \$x/\$y if the die roll is 1-3/4-6

V_A: value/token to A

V_B: value/token to B

p_a: price/apple

p_o: price/orange

TABLE 2. Experimental Tasks

We implemented our experiments on paper and followed up with some robustness treatments online (discussed further in Section 5 and Online Supplement I). In each session, paper instructions were provided and read aloud, subjects were given the opportunity to ask questions privately, and then participants completed a brief comprehension quiz and the experimenter individually checked answers and explained any errors. The comprehension quiz had each subject calculate how payment would be determined from their allocations when a decision involves two parts. This was designed to ensure that each subject was aware of how to calculate payments in these cases, but without instructing them to consider all possible combinations of allocations across parts.

Each decision had a cover page indicating the number of parts in the decision, with each part stapled beneath as a separate page. Thus, a subject was always informed when a decision contained multiple parts, but could choose whether or not to look at both parts before making allocations. A subject indicated each allocation by highlighting the line corresponding to their choice for each part using a provided highlighter. Subjects were allowed no other aids at their desk when making choices. Only one decision was handed out to a subject at a time, and that decision was collected before the next round was handed out. The order of decisions, and of parts within each decision, was varied across sessions to allow us to test and control for order effects.

Sessions took place in Toronto Experimental Economics Lab and SFU Experimental Economics Lab from June 2019 to February 2020, each taking place in a one hour block. Subjects were recruited from the labs' student participant pools to participate in one of the three experiments. Instructions, experimental materials, and details of the experimental procedure are provided in Online Supplement H.[19]

4. Results

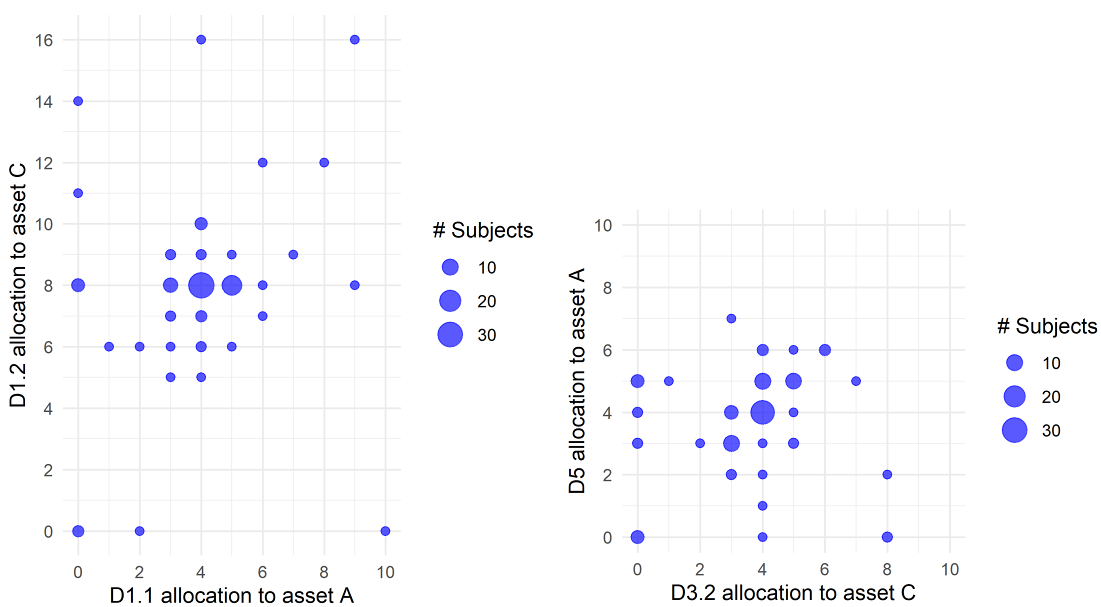
This section reports the results of our experimental tests of the models of bracketing. For each test, we also compute results allowing for one or two "errors" relative to its requirements. We define an error as how far we would need to move a subject's allocations for them to pass that test, measured in lines on the decision sheet(s).[20] For instance, in Risk and Social, a subject is within one error of passing a test if revising the choice by shifting one token from one asset/person to the other in a single part would lead to choices that pass that test. They are within two errors of passing a test if shifting two tokens from one asset/person to the other, either in the same part or in different parts, would lead to choices that pass. For predictions that require Walrasian budget sets, we modify the tests to account for discreteness in our experiment.

In Section 5, we examine the robustness of the following analysis to two particular concerns. First, we argue our results are robust to changes in the presentation of the decision. Second, broad bracketing requires allocating all the budget to one good in two-part decisions, and earlier work suggests subjects may be loath to do so. We argue there that neither affects our conclusions much.

[19]The three experiments were separately pre-registered through the Open Science Framework. Our analysis follows our plans with only minor modifications that do not affect the interpretation of our results. See Appendix G for details.
[20]Example decision sheets are provided in Figures 10, 11, 14, and 17 in Appendix H.

4.1. **Risk Experiment: How subjects behave.** Consider Decision 1 in the Risk Experiment. Since Asset B in the first part is a perfect substitute for Asset D in the second, comparative advantage requires that the subject first purchases it in the part where it has a lower opportunity cost. Consequently, a broad bracketer necessarily allocates all of their wealth in the first part to Asset B before they allocate any to Asset D in the second (as required by BB-Mon). If preferences are symmetric across the two equally-likely states, then broad bracketing further requires that $x_A^{1,1} = 0$. With risk aversion, broad bracketing makes even stronger predictions. The allocation $x_A^{1,1} = 0$ and $x_C^{1,2} = 14$ obtains a risk-free return of \$14, and any other feasible allocation results in a first- or second-order stochastically dominated distribution over returns at the decision-level. In contrast, a narrow bracketer facing Decision 1 allocates at least half of their budget to Asset B in Part 1. If they are also risk-averse, then they allocate exactly half their budget in Part 2 to each asset.

FIGURE 1. Allocations in Risk



(A) D1 Allocation in Risk (B) D3.2 and D5 Allocations

To compare how our subjects perform to this benchmark, Figure 1(A) plots the joint distribution of their allocations in D1. The x-coordinate describes their allocation to Asset A in D1.1, and the y-coordinate their allocation to Asset C in D1.2. The above discussion shows that broad bracketers will have allocations on the y-axis, and a risk-averse broad bracketer will select the allocation $\left(x_A^{1,1}, x_C^{1,2}\right) = (0, 14)$. Any narrow bracketer will select an allocation with $x_A^{1,1} \leq 5$, and any risk-averse narrow bracketer will select $x_C^{1,2} = 8$. The plot shows that few subjects are close to consistent with broad bracketing, and only a single subject makes an allocation close to the prediction of risk-averse broad bracketing. Narrow bracketing with risk aversion is consistent with the modal allocation of $\left(x_A^{1,1}, x_C^{1,2}\right) = (4, 8)$.

In our design, narrow bracketing makes strong predictions across decisions that have a part in common. For instance, D3.2 and D5 both ask subjects to allocate 10 tokens between identical assets with identical prices. A narrow bracketer would make the same allocation in each of the two parts. To illustrate, Figure 1(B) plots each subject's allocation in D3.2 against their allocation in D5. The x-coordinate describes their allocation to Asset C in D3.2, and the y-coordinate their allocation to its counterpart, Asset A in D5. A narrow bracketer's choices will fall on the 45-degree line as required by NB-WARP, and the farther the allocations are from that line the farther the subject is from narrow bracketing. In the plot, we can see that this prediction of narrow bracketing holds exactly for 54 of the 99 subjects. We visually represent all the data underlying our tests in Appendix B.

4.2. **Risk and Social Experiments: Revealed Preference Tests.** We begin by performing the direct revealed preference tests of bracketing developed in Section 1: NB-WARP, BB-WARP, and BB-Mon (Table 3).

Very few subjects are consistent with rationality and broad bracketing. There is only a single pair of decisions (D1 and D2) where choices could directly violate BB-WARP. For that pair, we test BB-WARP by comparing the final bundle for D1 to the final bundle for D2: for any choice of $x^{2,1}$ in D2 with $x_A^{2,1} \leq 8$, the same final bundle can be achieved in D1. In each of Risk and Social, only 20% of subjects are within one error of passing BB-WARP.[21] Even fewer subjects are consistent with BB-Mon than with BB-WARP. In Risk and Social respectively, we find that 8% and 12% of subjects are within one error of passing BB-Mon in

[21] Allowing for two errors substantially raises pass rates of BB-WARP in Risk and Social. As predicted by both risk averse narrow and broad bracketing, many subjects, 66% in Risk and 84% in Social, choose $x_A^{2,1} = 7$. The jump at two errors happens because a subject who allocates $x_A^{2,1} > 8$ trivially satisfies BB-WARP.

	Risk			Social		
# errors	0	1	2	0	1	2
NB-WARP (D1.1 and D5)	56	76	89	45	70	77
NB-WARP (D1.2 and D4))	56	74	81	63	78	82
NB-WARP (D3.2 and D5)	54	76	83	49	75	80
NB-WARP (D1.1 and D3.2)	49	76	85	51	83	87
NB-WARP (all)	29	44	61	28	54	64
BB-WARP (D1 and D2)	13	20	87	16	20	94
BB-Mon (D1)	12	13	15	14	14	14
BB-Mon (D3)	14	16	18	17	17	18
BB-Mon (both)	7	8	10	12	12	12
# subjects		99			102	

Entries count the # of subjects who pass test at the listed error allowance.

Table 3. Tests of NB-WARP and BB-WARP

both decisions. Looking separately at D1 and D3, between 13% - 17% of subjects are within one error of passing BB-Mon.

All told, the BB-WARP and BB-Mon tests provide evidence showing that 80%-92% of subjects are not broad bracketers. These rates of violations of broad bracketing are qualitatively similar to, but higher than, those found by Tversky & Kahneman (1981) (73%) and Rabin & Weizsäcker (2009) (28%-66%), and very close to the structural estimates of the latter (89%). In these prior experiments, each part consisted of a pairwise choice, so failures to broadly bracket are detected only for a particular range of risk preferences. In contrast, there are many ways a subject could reveal their failure to bracket broadly in our experiments, which gives us more power to detect failures.

While previous work can only falsify broad bracketing, our design allows us to test narrow bracketing as well. We test NB-WARP by comparing the allocations in each of the two parts that appear in multiple decisions. Specifically, NB-WARP requires that a subject makes the same choice in D1.1, D3.2, and D5, and the same choice in D1.2 and D4. Far more subjects pass each NB-WARP test than either BB-WARP or BB-Mon. Between 75-77% of subjects in Risk and 69-81% of subjects in Social are within one error of passing each of the pairwise NB-WARP tests. Allowing for one error, 44% and 53% of subjects pass all possible NB-WARP tests in Risk and Social, respectively.[22]

Result 1. *When allowing for one error, 44% and 53% of subjects are consistent with WARP under narrow bracketing, 20% and 20% are consistent with WARP under broad bracketing, and 8% and 12% are consistent with monotonicity under broad bracketing in the Risk and Social Experiments, respectively.*

We next conduct our revealed preference tests of the three models considered using the entire set of decisions for each subject. Notice that in both experiments, the alternative (x, y) leads to an identical outcome as the alternative (y, x).[23] As discussed in Section 1, we extend the tests so that whenever $(a, b)P^{\mathcal{D}}(x, y)$, i.e., (a, b) is directly revealed preferred to (x, y) in the dataset $\mathcal{D}$, we also have $(a, b)P^{\mathcal{D}}(y, x)$, as well as $(b, a)P^{\mathcal{D}}(x, y)$ and $(b, a)P^{\mathcal{D}}(y, x)$. This reduces the need to compare across decisions and makes our tests more demanding. For example, all three tests make point predictions in D2 and D4. Random behavior has a 0.001% chance or less of passing either BB- or NB-SARP with one error, and a 0.3% chance of passing the PNB Algorithm with one error (see Appendix D). Table 4 shows how many subjects pass each test.

	Risk			Social		
# errors	0	1	2	0	1	2
NB-SARP	23	34	43	15	36	44
BB-SARP	0	0	0	8	10	10
PNB	49	59	71	31	58	69
PNB-PE	51	63	75	39	67	76
# subjects		99			102	

Entries count the # of subjects who pass each test at the listed error allowance.

Table 4. Full Tests of Symmetric Models

[22] In contrast, randomly-generated choices have only a 0.4% chance of being within one error of passing all NB-WARP tests.

[23] Either leads to a lottery paying x and y with equal probability, or to a payment of x to one anonymous individual and y to another.

We compare results of the tests allowing for up to one error. We find that no subjects pass BB-SARP in Risk, and only 10% of subjects pass it in Social. However, 34-35% of subjects in each experiment pass NB-SARP. While fewer subjects pass BB-SARP than either BB-WARP or BB-Mon, a similar number of subjects pass all NB-WARP restrictions with one error and the more demanding NB-SARP with two errors. These results show that a plurality of our subjects are well-described as narrow bracketers.

Our test of partial-narrow bracketing diagnoses how many of those who fail BB-SARP and NB-SARP behave consistently with intermediate degrees of bracketing. We find that 15% of subjects in Risk and 12% of subjects in Social pass the PNB test but neither BB-SARP nor NB-SARP when allowing for one error.[24] Only 4% of subjects in Risk and 9% of subjects in Social pass the PNB-PE test but not the PNB test.

Result 2. *When allowing for one error relative to each test, 34% and 35% of subjects are rationalizable by narrow bracketing, 0% and 10% of subjects pass are rationalizable by broad bracketing in the Risk and Social Experiments, respectively. An additional 25% and 12% of subjects are rationalizable by partial narrow bracketing but not narrow nor broad bracketing.*

We note briefly that a subject whose underlying preferences are symmetric and linear would make the same choices regardless of how they bracket.[25] They would be indifferent between many bundles in our experiment and so violate our tests' assumption that choices reveal strict preferences. However, this cannot bias our conclusions too much since only five subjects in Risk and two subjects in Social could possibly be consistent with it.[26]

4.3. **Shopping Experiment: Induced-Value Tests.** The tests of our predictions thus far assume that utility is not observed. When utility is known, as in our Shopping Experiment,

[24]Broad and narrow bracketing are both special cases of partial-narrow bracketing, and thus any subject who passes BB- or NB-SARP will also pass this test.

[25]Symmetric linear preferences require only that $x_A^{1,1} = x_{A/C}^{3,2} = x_A^{5,1} = 0$. Linear preferences are the only class of preferences consistent with both narrow and broad bracketing (Appendix A, Theorem 4).

[26]One subject in each of Risk and Social pass NB-SARP, and the other subject in Social passes BB-SARP when allowing one error. Two other subjects in Risk also had $x_C^{1,2} = x_A^{3,1} = 0$, and so pass BB-WARP, BB-Mon, and NB-WARP but neither NB-SARP nor BB-SARP. In total, one in each of Risk and Social are classified as broad, two in Risk and one in Social are classified as narrow, and the remaining two in Risk are unclassified.

narrow and broad bracketing each make unique predictions in each decision.[27] To test the models, we compare how far each subject's choices are from each model's predictions (Table 5).

	D1			D3			Both			Full		
# errors	0	1	2	0	1	2	0	1	2	0	1	2
NB	23	23	60	53	65	69	20	21	49	15	16	40
BB	21	22	27	23	24	25	13	15	18	12	14	16
PNB	45	46	95	76	91	98	33	37	68	27	30	56
PNB-PE	45	46	95	76	93	98	33	37	70	27	30	57
# subjects	101											

Entries count the # of subjects who pass each test at the listed error allowance.

Table 5. Shopping Tests

Testing the point predictions of narrow bracketing in each of Decisions 1 and 3, 23% and 64% of subjects are respectively within one error of the predictions of narrow-bracketing, while 21% are consistent in both. Allowing for two errors raises pass rates to 59% for Decision 1 and 49% for both.[28] Looking at the full set of implications of narrow-bracketing on all choices made in the experiment, 40% of subjects are within two errors of passing.

In contrast, 21%, 24%, and 14% of subjects are within one error of being consistent with broadly-bracketed maximization in Decisions 1, 3, and both, respectively. When allowing for two errors, those numbers remain similar. Using all decisions in the experiment, only 16% of subjects are within two errors of being consistent with all implications of broad-bracketed maximization. PNB describes less than 2% of the remaining subjects.

Since we induce the payoff function, we are able to compute the value of α that best fits a subjects behavior (up to the limits imposed by discretization of the budget sets) under the assumption that the induced value function acts as their utility function. To

[27] Narrow-bracketed maximization implies $x_a^{1,1} = 1$, $x_a^{1,2} = 6$, $x_a^{3,1} = 5$, and $x_a^{3,2} = 4$. Broad-bracketed maximization implies $x_a^{1,1} = 0, x_a^{1,2} = 10, x_a^{3,1} = 10$, and $x_a^{3,2} = 0$.

[28] The sharp difference between one- and two-error tests in Decision 1 but not Decision 3 results from the fact that 40 subjects selected $x_a^{1,1} = 2, x_o^{1,1} = 4$, which is the second-best available bundle from a narrow bracketer's perspective but is two lines away from the best bundle of $x_a^{1,1} = 1, x_o^{1,1} = 6$ – due to the discreteness of the budget set, one apple and five oranges is between these bundles. We note that this deviation is in the opposite direction of that predicted by broad or partial-narrow bracketing, and 32 of these subjects also selected $x_a^{1,2} = x_o^{1,2} = 6$.

that end, we compute the point predictions of the partial-narrow bracketing model for each $\alpha \in \{0, 0.01, 0.02, \ldots, 0.99, 1\}$ for Decisions 1 and 3, and obtain distinct predictions for nine intervals of α. We assign each subject to the range of α for which their choices exhibit the fewest errors relative to that range's predictions. We find that 64% of subjects are classified to a range that includes full narrow bracketing, $\alpha = 1$, and 25% are classified to a range that includes full broad bracketing, $\alpha = 0$. Of the remaining subjects, none are best described by $\alpha \in [0.25, 0.71]$. This suggests that even those subjects who are not exactly described by either broad or narrow bracketing are close.[29]

4.4. **Classifying subjects to models.** The tests thus far do not make any adjustment for the fact that partial-narrow bracketing nests narrow and broad bracketing as polar cases, and can thus accommodate more behavior. To compare the predictive success of each model at the subject level, we use a subject-level implementation of the Selten score (Selten, 1991; Beatty & Crawford, 2011). For each subject and each model (symmetric versions for Risk and Social, using the induced value function for Shopping), we calculate the number of errors the subject exhibits relative to that model. Then, we calculate the number of possible choice combinations in the experiment that are consistent with that model and that number of errors – the model-error pair's "predictive area". We divide the predictive area by the total number of possible combinations of choices in the experiment to compute the measure for each subject i and model $m \in \{\text{broad, narrow, PNB, PNB-PE}\}$ as $\text{predictive\_success}_{i,m} = 1 - \frac{\#\text{predictive\_area\_for\_}i,m}{\#\text{all\_possible\_choices}}$.[30] We use all choices made in the experiment to assign each subject to the model with the highest predictive success; in cases where every rationalizing model-error pair for a subject would rationalize more than one million possible combinations of choices in our experiment, we categorize them as "Unclassified".

[29] Because of discreteness, $[0, 1]$ can be partitioned so that the parameters in each cell make the same choices. In this case, the partition is $\{[0, 0.04], [0.05, 0.06], [0.07, 0.14], [0.15, 0.24], [0.25, 0.29], [0.3, 0.38], [0.39, 0.71], [0.72, 0.75], [0.76, 1]\}$.

[30] In the Risk and Social Experiments, there are $(11 \times 17) \times 15 \times (11 \times 11) \times 17 \times 11 = 63,468,735$ possible combinations of choices. Symmetric narrow bracketing allows 6, 87, and 606 possible combinations of choices when allowing for zero, one, and two errors, respectively, whereas symmetric broad bracketing allows 12, 116, and 585 combinations of choices, symmetric PNB allows 35,797, 200,828, and 597,728 combinations, and symmetric PNB-PE allows 116,267, 619,375, and 1,725,466 combinations. Thus, a subject whose choices are consistent with partial-narrow bracketing will be classified as a partial-narrow bracketer if and only if they are sufficiently far from being consistent with both broad and narrow bracketing.

	Percent Selten Score Maximized		
	Risk	Social	Shopping
Broad Bracketing	2.02	9.8	26.73
	(0.55,7)	(5.52,17.46)	(19.26,36.46)
Narrow Bracketing	77.78	75.49	67.33
	(67.95,83.99)	(67.64,84.47)	(58.27,76.45)
PNB	7.07	1.96	3.96
	(3.43,13.74)	(0.55,7.01)	(1.57,9.84)
PNB-PE	0	3.92	0.99
	(0,3.7)	(1.57,9.85)	(0.18,5.45)
Unclassified	13.13	8.82	0.99
	(7.76,20.97)	(4.81,16.24)	(0.18,5.45)

To calculate confidence intervals, we assume that the model of bracketing is multinomially distributed with fixed strike rate within each treatment and calculate the ? score.

Table 6. Classification of subjects

Across the three experiments, we classify 67-78% of subjects as narrow bracketers (Table 6). In contrast, 2%, 10%, and 27% of subjects are classified as broad bracketers in the Risk, Social, and Shopping Experiments respectively. However, 7%, 6%, and 5% were classified to one of the two partial-narrow bracketing models. After adjusting for predictive power, partial-narrow bracketing does not help explain very many subjects' behavior.

To form confidence intervals, we assume that model of bracketing is multinomially distributed with fixed strike rate within each treatment. We calculate confidence intervals according to the Wilson (1927a) score. In all three treatments, the lower bound of the interval for narrow bracketers exceeds the upper bound of the intervals for all other models of bracketing. Most other intervals overlap, but there are significantly more broad bracketers than partial-narrow bracketers in the Shopping Experiment.

Result 3. *Judging each model's fit by its predictive success, 67-78% of subjects are classified as narrow bracketers, 2-27% are classified as broad bracketers, and 5-7% are classified as partial-narrow bracketers across the three experiments.*

5. Robustness and Secondary Analyses

To study how choice architecture mediates bracketing, we conducted an online version of our Risk Experiment that varied the presentation in a two-by-two design. First, the "Examine" treatment instructed the subject to "First, examine both accounts, then purchase your investments" in the instructions, tested this in a quiz question, then included that text at the top of each two-part decision screen; the "Basic" treatment did not. Second, the "Tabs" treatment presented each part of a decision as a separate HTML tab (analogous to our separate pages in our paper experiments), whereas the "Side-by-Side" treatment presented parts of a decision side-by-side on the same screen (analogous to Tversky & Kahneman 1981). We recruited 200 US-based subjects from Prolific Academic to participate and randomly assigned each to one of the four treatment pairs.

We also replicated our Shopping experiment online with 46 subjects from Prolific Academic. For all 46, we implemented the Side-by-Side and Examine interventions, eliminated the quantity-restricted sale in D2, and provided subjects with a calculator instead of a payoff table. We provide detailed screenshots and results in the Online Supplement.

In Section 5.1, we use the online experiments to argue that presentation has a limited effect on our results.The online experiments allow us to collect non-choice data that shed light on the subjects' choice processes. In Section 5.4, we use this data to argue that a substantial fraction of narrow bracketers give some consideration to both parts of the decision before choosing. We then explore the robustness of our tests of broad bracketing to extremeness aversion, investigate the possibility of order effects or learning, and perform statistical tests of the differences we find across the domains.

5.1. **Effects of choice architecture.** The online subject pool is more slanted towards narrow bracketing than our original Pen-and-paper study. Pass rates for each NB-WARP test are higher, and pass rates for BB-WARP and BB-Mon are generally lower. Only 2 of 200 (1%) subjects pass BB-SARP when allowing for two errors, and these two, and only these two, are classified as broad bracketers according to Selten score. In contrast, 96 (48%) subjects pass NB-SARP when allowing one error, and 84% are classified as narrow bracketers by Selten score.

We find almost no effect on choices from either of the two interventions in the Online Risk Experiment. Both subjects who pass BB-SARP are in the Examine and Tabs treatment, contrary to our expectation that Side-by-Side would be more conducive to broad bracketing. Neither treatment has a large or statistically significant (with a small sample size caveat) effect on the rate of broad bracketing ($p = 0.22, p = 0.24$, for Fisher's exact tests of the Examine and Tabs treatments, respectively). Within the Tabs group, Examine does not have a statistically significant effect ($p = 0.20$, Fisher's exact test). We similarly find no effect of the treatments on the rates of narrow bracketing ($p = 1.00$ for both, Fisher's exact tests).

In addition, the Online Shopping Experiment implemented both Examine and Side-by-Side. It provided subjects with a calculator instead of a payoff table. We classified 5 out of 46 (10.86%) subjects to broad bracketing and 33 (71.74%) to narrow. The rate of narrow bracketing did not change much, while the rate of broad bracketing slightly decreased relative to the pen-and-paper Shopping Experiment ($p = 0.03$, Fisher's exact test).

All-in-all, this suggests that the low rates of broad bracketing we find are not overly sensitive to varying the choice architecture to encourage broad bracketing. We caution against over-interpreting this result since broad bracketing was so rare in all treatments. The shift to an online interface and the Prolific subject pool may have had more effect than any of the nudges. We suspect that more extreme nudges and decision aids might be more effective.

5.2. **Non-choice data and bracketing.** Our online experiments shed light on how the choice process, and in particular consideration, differed across subjects. We have two tools for measuring consideration. In the Tabs arm of Online Risk, we observe which tabs subjects clicked and when they made their decision. In Online Shopping, we record how subjects used the calculator. This non-choice data shows that only some narrow bracketers completely ignore other parts of the decision. In both experiments, more than a quarter of narrow bracketers gave some consideration to both parts of the decision. These subjects had enough information to bracket more broadly yet did not.

In the Tabs versions of the Online Risk Experiment, we record whether each subject clicked on both tabs before making their final choices. Only 28 of 102 subjects did so in both D1 and D3. While this includes the 2 broad bracketers, the other 26 subjects had sufficient information to bracket broadly but failed to do so. Of the 28, 22 were in the Examine-Tabs treatment, and 6 were in the Basic-Tabs treatment ($p < 0.01$, Fisher's exact). The prompt was effective at increasing this click pattern, but it did not significantly affect the classification to broad bracketing ($p = 0.20$, Fisher's exact). Clearly, paying attention to both parts of a decision is necessary to bracket broadly. However, the fraction of narrow bracketers among those who paid attention was not substantially different than the fraction among those who did not (21 of 28 versus 61 of 74, $p = 0.41$, Fisher's exact). At the other extreme, we find that 33 subjects made their final choice in each of D1 and D3 before having seen both parts, and thus could not have been broad bracketing; 29 of these subjects are classified as narrow bracketers.

While intensity of calculator use does not differ much between broad and narrow bracketers in the Online Shopping Experiment, patterns of calculator use do.[31] Ex-ante, we expected that plugging in bundles that are feasible in the decision as a whole (e.g. (9, 11) for D3) to the calculator (a "broad calculation") would predict broad bracketing, and that plugging in a bundle that is available in a part of a decision but lies below the decision-level feasible set (e.g. (5,5) for D3) to the calculator (a "narrow calculation") would predict narrow bracketing. We classify nine of ten subjects of who made a narrow calculation as narrow bracketers. Surprisingly, we classify nine of 16 subjects who made at least one such "broad calculation" as narrow bracketers, and only three as broad. These subjects appear to actively attend to how the parts fit together, yet still make narrowly-bracketed choices.

In summary, the non-choice data suggest diversity in the relationship between consideration and bracketing. It suggests that about a quarter of narrow bracketers appear to follow a choice process that involves only considering one part at a time, but this same data also shows that a quarter or more of narrow bracketers gave some consideration to both parts.

[31]The average (median) number of calculations was 10.40 (12) for broad bracketers, 12.21 (8) for narrow bracketers, and 12.86 (10) for all subjects. As a whole,

The seemingly weak link between broader consideration and broadly-bracketed choices suggests why the interventions we designed to encourage broad bracketing in the Online Risk Experiment were not very effective. However, the sparsity of broad bracketers in both experiments makes it hard for us to tell what drives it. The choice processes leading to narrow or broad bracketing is a fertile direction for future work.

5.3. **Effects of Extremeness Aversion.** As we describe in Section 1, broad bracketing requires a corner solution in at least one part of any multi-part decision. Evidence from other contexts suggests that some subjects may be *extremeness averse* and avoid corner choices. For example, subjects in linear public good games tend not to play the Nash strategy of making no contributions. However, in non-linear public good games where the Nash strategy requires a positive but not complete contribution, subjects play the equilibrium strategy more frequently (see e.g. Section 2 of Vesterlund, 2016, for a discussion). Extremeness aversion could affect our conclusions about broad bracketing and its prevalence relative to narrow bracketing.

Non-extreme allocations in two-part decisions lead to violations of BB-Mon. However, few additional subjects are consistent with a relaxation of BB-Mon that allows for subjects to be close to, but not necessarily at, a corner when required. Formally, we say that a subject passes *Extremeness-Averse (EA-)BB-Mon* if they are within 2 tokens of making an extreme allocations in each decision required by BB-Mon. Without allowing for any errors, 3 subjects in Risk and 0 subjects in Social pass EA-BB-Mon but not BB-Mon. Allowing for 2 errors, 14% of subjects in Risk and 3% in Social pass EA-BB-Mon but not BB-Mon. As a benchmark, 26% of all possible datasets pass EA-BB-Mon but not BB-Mon when allowing for two errors.

Extremeness aversion is unlikely to affect our classification or tests. Of the 77 subjects classified as narrow bracketers in each of Risk and Social, none pass EA-BB-Mon but not BB-Mon (allowing for two errors, this increases to six and one, respectively). No subject in either Risk or Social passes both NB-SARP and EA-BB-Mon but not BB-Mon when allowing for no or one errors in each test. Allowing two errors in each test, only 3 subjects in Risk and 1 in Social pass both NB-SARP and EA-BB-Mon but not BB-Mon. Since BB-Mon is a necessary but not sufficient condition for broad bracketing, a highly risk-tolerant or inequity-neutral narrow bracketer would pass both NB-SARP and BB-Mon. Only one subject in each does so. Somewhat more risk- or inequity-averse narrow bracketers will also pass EA-BB-Mon.

This suggests that the effect of extremeness aversion on our tests is probably not too large. In addition, a similar number of subjects are consistent with broad bracketing in decisions that require two corner choices as are consistent in decisions that require only one corner choice.[32] The evidence still suggests heterogeneity in bracketing, with a plurality best described as narrow bracketers, even after adjusting for extremeness aversion.[33]

5.4. **Welfare losses and measurement.** So far, we have focused on identifying how our subjects bracketed. We now apply our classification to quantify in dollar terms the payoffs forgone by narrow bracketers (Table 7). We also illustrate how looking at aggregate data can lead to misleading inferences. A narrow bracketer's preference parameters should be measured using parts as the unit of observation, and broad bracketer's should use decisions as the unit of observation. Using the wrong unit of observation, necessary for an aggregate analysis with heterogeneous bracketing, leads to misleading conclusions about preferences. We focus on the Social and Shopping Experiments because there are only two participants classified as broad bracketers in Risk.

In the Shopping Experiment, we can precisely measure welfare losses through the earnings that participants did not receive.[34] Narrow bracketers had lower earnings than broad bracketers in both two-part decisions. They had \$1.41 and \$1.67 in forgone payments compared

[32] A broad bracketer in Shopping buys only apples in D3.1, only oranges in D3.2, and only oranges in D1.1, but buys both fruits in D1.2. D3 requires two extreme choices, and D1 requires only one. As reported in Table 5, there are a very similar number of subjects close to these predictions in D1 and D3.

[33] We also conducted a follow-up experiment for which only one of the two treatments requires a corner choice. To get more separation between narrow and broad, we used a much finer choice grid for each budget. Unfortunately, this made the data too noisy to draw strong conclusions. No more than 14.8% of subjects are close to the predictions of either narrow or broad bracketing combined in either treatment. Nonetheless, the ratio of broad to narrow bracketers does not vary much across treatments, lending weak support to our conclusion. We elaborate on this experiment in the Online Supplement, Section J.

[34] Interpreting payoff losses as welfare losses requires that broad bracketing does not have a direct utility cost.

Panel A

	Payoff loss ($)							
	D1.1	D1.2	D1	D3.1	D3.2	D3	D5	n
Risk	0.39	0	0.39	0	0.35	0.35	0.38	99
NB	0.4	0	0.4	0	0.38	0.38	0.41	77
BB	0	0	0	0	0.2	0.2	0.3	2
Social	0.47	0	0.47	0	0.46	0.46	0.49	102
NB	0.51	0	0.51	0	0.51	0.51	0.49	77
BB	0	0	0	0	0.03	0.03	0.44	10
Shopping	0.56	0.51	1.06	1.03	0.85	1.25	0.05	101
NB	0.22	0.07	1.41	0.01	0.07	1.67	0.03	68
BB	1.23	1.31	0.35	3.76	2.94	0.14	0.04	27

"Payoff loss" columns report total $ loss (Shopping), loss in expected value (Risk), and average loss per recipient (Social) compared to expected/average value maximization. We evaluate each part on its own in columns D1.1 and D3.2 and integrate across parts in columns D1 and D3. In D1.2 and D3.1 all allocations have no loss in either Risk or Social.

Panel B

	Payoff difference ($)							
	D1.1	D1.2	D1	D3.1	D3.2	D3	D5	n
Risk	4.24	2.24	5.39	1.88	4.7	4.51	3.79	99
NB	3.47	0.91	3.91	0.7	3.72	3.98	3.24	77
BB	12	9	3	3	7.6	4.6	6.6	2
Social	3.1	2.39	1.54	2.29	3.35	1.91	2.58	102
NB	1.64	0.52	1.62	0.6	1.68	1.68	2.17	77
BB	12	12	0	10	11.34	1.46	2.56	10

"Payoff difference" columns report the difference in $ allocations across the two states or people.

TABLE 7. Losses and payoff differences by parts and decisions

to just $0.35 and $0.14. The differences between the maximum and the minimum payoffs are $9.60 and $10.36 in D1 and D3, so the payoff losses for narrow bracketers are 10.9% and 14.8% of the variable payment. In the one-part decision D5, narrow and broad bracketers should make identical choices. They do indeed make very similar choices, suggesting that arithmetic errors do not explain the difference. Indeed, narrow bracketers do a good job at optimizing part-by-part, and would forgo no more than $0.22 if the compensation was based on individual parts of a decision rather than the decision as a whole.[35]

[35] There is a similar pattern for expected or aggregate payoffs in the Risk and Social Experiments, but risk or inequity aversion make interpreting this as a welfare loss more ambiguous.

The "Payoff difference" columns of Table 7 illustrate how it is easy to obtain misleading inferences about how subjects make the equity-efficiency trade-off. In the one-part decision D5, narrow bracketers implemented allocations that differed by an average of \$2.17 between A and B, while the corresponding average difference is \$2.56 for broad bracketers ($p = 0.18$, rank sum test comparing D5 allocations). Narrow and broad bracketers make similar trade-offs between equity and efficiency from one-part decisions. Yet if we look at D1.2 or D3.1 on its own, ignoring the other part of the decision, we would incorrectly infer that broad bracketers care mainly about efficiency, while narrow bracketers make near-equal allocations on average, consistent with inequity aversion. We would obtain an opposite conclusion by looking at D1 as a whole, where narrow bracketers' decision-level allocations involve some inequity on average, but all broad bracketers make perfectly equal final allocations.

5.5. **Differences across experiments.** The fraction of subjects classified to broad bracketing varies across experiments, from 1% for Online Risk to 27% in Pen-and-paper Shopping. There is a significant difference in broad bracketing rates between Pen-and-paper Shopping and each of Risk and Social ($p < 0.01$, Fisher's exact test) as well as between the Online and Pen-and-paper Shopping experiments ($p = 0.03$).[36] However, narrow bracketing rates varied much less, from 67% to 78% across Pen-and-paper experiments, with no significant differences in these rates ($p > 0.10$ for all pairwise Fisher's exact tests). Nor was there a significant difference in narrow bracketing rates between the Online and Pen-and-paper versions of Risk and Shopping ($p > 0.10$ for both Fisher's exact tests). Explanations for the higher rate of broad bracketing in Shopping and lower rate in Risk include the more naturalistic setting, the presence of an objectively-correct payoff function, and cognitive difficulties specific to choice under risk (as suggested by Martínez-Marquina *et al.* 2019). We cannot distinguish between these explanations.

6. Conclusion

We propose revealed preference tests for how a person brackets their choices that rely only on monotonicity of underlying preferences. We deploy these tests in an experiment

[36]The difference between Online and Pen-and-paper Risk was not significant, but this is to be expected given the negative effect on broad bracketing and the already small numbers of broad bracketers.

where both narrow and bracketing make falsifiable predictions, unlike in past work. Across our experiments, at least twice as many subjects were classified as narrow bracketers than as broad bracketers. A majority of people tend to narrowly bracket, while a noticeable minority broadly bracket. While many of our subjects are not well-described by either broad or narrow bracketing, our novel tests of partial-narrow bracketing suggest that it does not do much better after adjusting for predictive power. This suggests that applications should calibrate a population mix of broad and narrow bracketers rather than a representative agent model with a calibrated partial-narrow bracketing parameter (as in Barberis & Huang, 2007).

Bracketing rates differ across tasks, domains, and subject pools. While our framework is well-suited to detect and to measure these differences, it is less well-suited to determine why these differences persist. Non-choice data, as we collected in the online follow up experiment, can help. It seems to rule out uniform explanations for why so many bracket narrowly, such as lack of awareness of complementarities across parts. We think understanding why people bracket the way they do is an interesting direction for future work.

References

Abeler, Johannes, & Marklein, Felix. 2017. Fungibility, labels, and consumption. *Journal of the European Economic Association*, **15**(1), 99–127.

Aczél, J. 1966. *Lectures on Functional Equations and Their Applications.* New York and London: Academic Press.

Andreoni, James, & Miller, John. 2002. Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica*, **70**(2), 737–753.

Baillon, Aurélien, Halevy, Yoram, & Li, Chen. 2022. Randomize at your own Risk: on the Observability of Ambiguity Aversion. *Econometrica*, **90**(3), 1085–1107.

Barberis, N., & Huang, M. 2007. The loss aversion/narrow framing approach to the equity premium puzzle. *Pages 199–229 of:* Mehra, R (ed), *Handbook of Investments: Equity Premium.* Amsterdam: North-Holland.

Barberis, Nicholas. 2012. A model of casino gambling. *Management Science*, **58**(1), 35–51.

Barberis, Nicholas, & Huang, Ming. 2009. Preferences with frames: a new utility specification that allows for the framing of risks. *Journal of Economic Dynamics and Control*, **33**(8), 1555–1576.

Barberis, Nicholas, Huang, Ming, & Thaler, Richard H. 2006. Individual preferences, monetary gambles, and stock market participation: A case for narrow framing. *American Economic Review*, **96**(4), 1069–1090.

Baron, Jonathan, & Kemp, Simon. 2004. Support for trade restrictions, attitudes, and understanding of comparative advantage. *Journal of Economic Psychology*, **25**(5), 565–580.

Beatty, Timothy KM, & Crawford, Ian A. 2011. How demanding is the revealed preference approach to demand? *American Economic Review*, **101**(6), 2782–95.

Benartzi, Shlomo, & Thaler, Richard H. 1995. Myopic loss aversion and the equity premium puzzle. *Quarterly Journal of Economics*, **110**(1), 73–92.

Blow, Laura, & Crawford, Ian. 2018. *Observable Consequences of Mental Accounting*. Working Paper.

Bolton, Gary E, & Ockenfels, Axel. 2000. ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, **90**(1), 166–193.

Bronars, Stephen G. 1987. The power of nonparametric tests of preference maximization. *Econometrica*, **55**(3), 693–698.

Camerer, Colin F. 2004. Prospect Theory in the Wild: Evidence from the Field. *Pages 148–162 of:* Camerer, Colin F, Loewenstein, George, & Rabin, Matthew (eds), *Advances in Behavioral Economics*. Princeton University Press.

Charness, Gary, & Rabin, Matthew. 2002. Understanding social preferences with simple tests. *Quarterly Journal of Economics*, **117**(3), 817–869.

Chen, Daniel L, Schonger, Martin, & Wickens, Chris. 2016. oTree – An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, **9**, 88–97.

Choi, Syngjoo, Fisman, Raymond, Gale, Douglas, & Kariv, Shachar. 2007. Consistency and heterogeneity of individual behavior under uncertainty. *American Economic Review*,

97(5), 1921–1938.

Clark, Stephen A. 1993. Revealed preference and linear utility. *Theory and Decision*, **34**(1), 21–45.

Cox, James C, Sadiraj, Vjollca, & Schmidt, Ulrich. 2015. Paradoxes and mechanisms for choice under risk. *Experimental Economics*, **18**(2), 215–250.

Dhami, Sanjit. 2016. *The foundations of behavioral economic analysis*. Oxford University Press.

Ebert, Sebastian, & Strack, Philipp. 2015. Until the bitter end: on prospect theory in a dynamic context. *American Economic Review*, **105**(4), 1618–33.

Ellis, Andrew, & Piccione, Michele. 2017. Correlation Misperception in Choice. *American Economic Review*, **107**(4), 1264–92.

Evers, Ellen, Imas, Alex, & Kang, Chisty. 2022. *Mental accounting, similarity, and preferences over the timing of outcomes*. Tech. rept. 4.

Exley, Christine L, & Kessler, Judd B. 2018. *Equity concerns are narrowly framed*. Tech. rept. National Bureau of Economic Research.

Eyster, Erik, & Weizsäcker, Georg. 2016. Correlation neglect in portfolio choice: Lab evidence. *SSRN Working Paper 2914526*.

Fallucchi, Francesco, & Kaufmann, Marc. 2021. Narrow bracketing in work choices. *arXiv preprint arXiv:2101.04529*.

Fisman, Raymond, Kariv, Shachar, & Markovits, Daniel. 2007. Individual preferences for giving. *American Economic Review*, **97**(5), 1858–1876.

Galperti, Simone. 2019. A theory of personal budgeting. *Theoretical Economics*, **14**(1), 173–210.

Gneezy, Uri, & Potters, Jan. 1997. An experiment on risk taking and evaluation periods. *Quarterly Journal of Economics*, **112**(2), 631–645.

Hastings, Justine, & Shapiro, Jesse M. 2018. How are SNAP benefits spent? Evidence from a retail panel. *American Economic Review*, **108**(12), 3493–3540.

Hastings, Justine S, & Shapiro, Jesse M. 2013. Fungibility and consumer choice: Evidence from commodity price shocks. *Quarterly Journal of Economics*, **128**(4), 1449–1498.

Heimer, Rawley, Iliewa, Zwetelina, Imas, Alex, & Weber, Martin. 2020. *Dynamic Inconsistency in Risky Choice: Evidence from the Lab and Field*. Tech. rept. SSRN Working Paper 3600583.

Hines, James R, & Thaler, Richard H. 1995. The flypaper effect. *Journal of Economic Perspectives*, **9**(4), 217–226.

Imas, Alex. 2016. The realization effect: Risk-taking after realized versus paper losses. *American Economic Review*, **106**(8), 2086–2109.

Kahneman, Daniel, & Tversky, Amos. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, **47**(2), 263–291.

Kassambara, Alboukadel. 2020. *ggpubr: 'ggplot2' Based Publication Ready Plots*. R package version 0.4.0.

Koch, Alexander K, & Nafziger, Julia. 2019. Correlates of narrow bracketing. *Scandinavian Journal of Economics*, **121**(4), 1441–1472.

Kőszegi, Botond, & Matějka, Filip. 2020. Choice simplification: A theory of mental budgeting and naive diversification. *Quarterly Journal of Economics*, **135**(2), 1153–1207.

Kőszegi, Botond, & Rabin, Matthew. 2006. A model of reference-dependent preferences. *Quarterly Journal of Economics*, **121**(4), 1133–1165.

Levin, Vladimir L. 1983. Measurable Utility Theorems for Closed and Lexicographic Preference Relations. *Soviet Mathematics Doklady*, **27**(3), 639–43.

Lian, Chen. 2019. *A theory of narrow thinking*. Tech. rept. MIT mimeo. Working Paper.

Martínez-Marquina, Alejandro, Niederle, Muriel, & Vespa, Emanuel. 2019. Failures in contingent reasoning: The role of uncertainty. *American Economic Review*, **109**(10), 3437–3474.

Mas-Colell, Andreu, Whinston, Michael D., & Green, Jerry R. 1995. *Microeconomic Theory*. New York: Oxford University Press.

Mu, Xiaosheng, Pomatto, Luciano, Strack, Philipp, & Tamuz, Omer. 2020. Background risk and small-stakes risk aversion. *arXiv preprint arXiv:2010.08033*.

Nishimura, Hiroki, Ok, Efe A., & Quah, John K.-H. 2017. A Comprehensive Approach to Revealed Preference Theory. *American Economic Review*, **107**(4), 1239–63.

Odean, Terrance. 1998. Are investors reluctant to realize their losses? *Journal of Finance*, **53**(5), 1775–1798.

O'Donoghue, Ted, & Sprenger, Charles. 2018. Reference-dependent preferences. *Pages 1–77 of: Handbook of Behavioral Economics: Applications and Foundations 1*, vol. 1. Elsevier.

Polisson, Matthew, Quah, John K.-H., & Renou, Ludovic. 2020. Revealed Preferences over Risk and Uncertainty. *American Economic Review*, **110**(6), 1782–1820.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rabin, Matthew. 2000. Risk Aversion and Expected-Utility Theory: A Calibration Theorem. *Econometrica*, **68**(5), 1281–92.

Rabin, Matthew, & Weizsäcker, Georg. 2009. Narrow bracketing and dominated choices. *American Economic Review*, **99**(4), 1508–43.

Read, Daniel, Loewenstein, George, & Rabin, Matthew. 1999. Choice Bracketing. *Journal of Risk & Uncertainty*, **19**(1-3), 171–197.

Safra, Zvi, & Segal, Uzi. 2008. Calibration results for non-expected utility theories. *Econometrica*, **76**(5), 1143–1166.

Samuelson, Paul. 1963. Risk and uncertainty: A fallacy of large numbers. *Scientia*, **57**(98).

Selten, Reinhard. 1991. Properties of a measure of predictive success. *Mathematical Social Sciences*, **21**(2), 153–167.

Shefrin, Hersh, & Statman, Meir. 1985. The disposition to sell winners too early and ride losers too long: Theory and evidence. *Journal of Finance*, **40**(3), 777–790.

Smith, Adam. 1776 (1904 edition). *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: Methuen & Co.

Snowberg, Erik, & Yariv, Leeat. 2021. Testing the waters: Behavior across participant pools. *American Economic Review*, **111**(2), 687–719.

Sobel, Joel. 2005. Interdependent preferences and reciprocity. *Journal of Economic Literature*, **43**(2), 392–436.

Thaler, Richard H. 1985. Mental accounting and consumer choice. *Marketing Science*, **4**(3), 199–214.

Thaler, Richard H. 1999. Mental accounting matters. *Journal of Behavioral Decision Making*, **12**(3), 183–206.

Thaler, Richard H, & Johnson, Eric J. 1990. Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. *Management Science*, **36**(6), 643–660.

Tversky, Amos, & Kahneman, Daniel. 1981. The framing of decisions and the psychology of choice. *Science*, **211**(4481), 453–458.

Vesterlund, Lise. 2016. Using experimental methods to understand why and how we give to charity. *Chap. 2, pages 91–151 of:* Kagel, John H., & Roth, Alvin E. (eds), *Handbook of Experimental Economics*, vol. 2. Princeton, NJ, USA: Princeton University Press.

Vorjohann, Pauline. 2020. *Reference-dependent choice bracketing.* Tech. rept. University of Exeter. Working Paper.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York.

Wickham, Hadley, & Henry, Lionel. 2020. *tidyr: Tidy Messy Data.* R package version 1.1.0.

Wilson, E. B. 1927a. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, **22**(158), 209–212.

Wilson, Edwin B. 1927b. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, **22**(158), 209–212.

Zhang, Mu. 2023. *Procedural Expected Utility.* Tech. rept. Working Paper.

Appendix A. Theoretical Appendix: Proofs, Derivations, and PNB Algorithm

Proof of Theorem 1. One can follow the usual proof that SARP holds if and only if a dataset is rationalizable by a complete and transitive preference relation to establish the result with a preference relation; see e.g. Mas-Colell *et al.*, 1995. Establishing rationalization with a utility function requires a bit more work. We provide an argument below as we are unaware of one in the literature.

Given $\mathcal{O} = (x^i, B^i)_{i=1}^N$ for either $\mathcal{O} = \mathcal{D}^{NB}$ or $\mathcal{O} = \mathcal{D}^{BB}$, set $X' = \cup_i B^i$. Observe that each B^i is a compact set, either by assumption (for $\mathcal{O} = \mathcal{D}^{NB}$) or since the finite sum of compact sets is compact (for $\mathcal{O} = \mathcal{D}^{BB}$). Thus X' is also compact. It is therefore bounded as a subset of $\mathbb{R}^n$. Let X be a closed ball centered at 0 containing X'.

Let $\succsim$ be the transitive closure of $P^{\mathcal{O}}$ defined on X. We show that $\succsim$ is closed as a subset of $X \times X$. First, note $P^{\mathcal{O}}$ itself is closed as a finite union of the closed sets $\{(y, z) \in X \times X : y \geq z\}$ and $\{x^i\} \times B^i$ for all $i = 1, \ldots, N$. Second, we show that $y \succsim z$ if and only if there exist $y_1, y_2, \ldots, y_{K-1}, y_K$ so that $y_i P^{\mathcal{O}} y_{i+1}$ for all i, $y_1 = y$, and $y_K = z$ where $K \leq 2N + 1$. Consider any $y \succsim z$. By definition, there exist $y_1, y_2, \ldots, y_K$ so that $y_i P^{\mathcal{O}} y_{i+1}$ for all i, $y_1 = y$, and $y_K = z$ and there is no shorter sequence that establishes $y \succsim z$. If $y_j \geq y_{j+1}$ and $y_{j+1} \geq y_{j+2}$ for some i, then $y_j \geq y_{j+2}$, and we can construct a shorter sequence by leaving y_{j+1} out. If $y_j \not\geq y_{j+1}$, then $y_j = x^{i(j)}$ and $y_{j+1} \in B^{i(j)}$ for some $i(j)$. Then, $y_k \neq y_j$ for all $k > j$ by SARP, so at most one index j' has $i(j') = i$ for each i. Therefore, the longest possible shortest sequence alternates $\geq$ with direct revelations from choice. There can be no more than $2N + 1$ of these, so $K \leq 2N + 1$. Moreover, since we can trivially append $y_{K+1} = y_K$ to the end of such sequences, we can consider only sequences with exactly $2N+1$ in what follows.

Now, pick any sequence $(y_m, z_m)_{m=1}^{\infty}$ contained in $\succsim$ that converges to (y, z). Then, there exists $y_m^2, \ldots, y_m^{2N}$ so that $y_m P^{\mathcal{O}} y_m^2 P^{\mathcal{O}} \ldots P^{\mathcal{O}} y_m^{2N} P^{\mathcal{O}} z_m$. The sequence $(y_m, y_m^2, \ldots, y_m^{2N}, z_m)$ has a convergent subsequence $(y_{m_k}, y_{m_k}^2, \ldots, y_{m_k}^{2N}, z_{m_k})$ since $X \times \cdots \times X$ is compact by the Tychonoff Theorem. Let $(y^*, y^{2*}, \ldots, y^{2N*}, z^*)$ be its limit. Since $P^{\mathcal{O}}$ is closed, $y^* P^{\mathcal{O}} y^{2*} P^{\mathcal{O}} \ldots P^{\mathcal{O}} y^{2N*} P^{\mathcal{O}} z^*$. Finally, $(y^*, z^*) = (y, z)$ since $(y_m, z_m) \to (y, z)$. Hence $\succsim$ is closed.

Now, apply Levin, 1983's Theorem, as in Nishimura *et al.*, 2017, to get a utility function U from X to $[0, 1]$ so that $x \succ y$ implies $U(x) > U(y)$ and $x \succsim y$ implies $U(x) \geq U(y)$. Since $\geq$ is included in $\succsim$, $U(\cdot)$ is increasing. Since $x^i \succ y$ for all $y \in B^i \setminus \{x^i\}$, $U(x^i) > U(y)$ for all $y \in B^i \setminus \{x^i\}$. Extending to all of $\mathbb{R}_+^n$ is straightforward, simply set $U(y) = V(y)$ whenever $y \notin X$ for an arbitrary increasing function $V(\cdot)$ picked so that $V(y) > 1$ for all y.

Proof of Theorem2. Fix α and $\mathcal{D}$. For every decision t, define the lottery

$$p^t = \frac{\alpha K_t}{\alpha K_t + (1-\alpha)} \left(\frac{1}{K_t}, x^{t,k}\right)_{k=1}^{K_t} + \frac{(1-\alpha)}{\alpha K_t + (1-\alpha)} \left(1, x^t\right),$$

and $Y \subseteq \mathbb{R}^n$ to be a finite set that includes the union of the supports of $p^1, \ldots, p^T$ and the vector 0. It will be convenient to take Y equal to this set but this is inessential. Denote by ΔY the finite support lotteries over Y. Consider the set

$$Q_Y^t = \left\{ \frac{\alpha K_t}{\alpha K_t + (1-\alpha)} \left(\frac{1}{K_t}, y^k\right)_{k=1}^{K_t} + \frac{(1-\alpha)}{\alpha K_t + (1-\alpha)} (1, y) \in \Delta Y : \exists z^k \in B^{t,k} \text{ so that } y^k \leq z^k \,\&\, y \leq \sum z^k \right.$$

is a finite set of lotteries over Y that includes p^t and any others that are affordable. The ancillary dataset

$$\mathcal{D}_Y^\alpha = \left\{ \left(p^t, Q_Y^t\right) \right\}_{t=1}^T$$

is strictly rationalizable by expected utility if and only if $\mathcal{D}$ is rationalizable by α-partial-narrow bracketing.

If so, then there exists an EU preference V over ΔY so that

$$V\left(p^t\right) > V\left(q^t\right) \; \forall q^t \in Q_Y^t \backslash p^t$$

for all t. Let v be the utility index of V and extend to $\mathbb{R}_+^n$ by

$$u^*(z) = \max_{y \leq z, y \in Y} v(y).$$

Then for any t and $y^{t,1} \times \cdots \times y^{t,K_t} \in B^{t,1} \times \cdots \times B^{t,K_t} \setminus \left\{x^{t,1} \times \cdots \times x^{t,K_t}\right\}$,

$$\alpha \sum_k u^*(x^{t,k}) + (1-\alpha) u^*(x^t) = \kappa_t V\left(p^t\right) > \kappa_t \max_{q^t \in Q_Y^t \backslash p^t} V\left(q^t\right) \geq \alpha \sum_k u^*(y^{t,k}) + (1-\alpha) u^*\left(\sum_{k=1}^{K_t} y^{t,k}\right)$$

where $\kappa_t = \alpha K_t + (1-\alpha)$.[37] To find a strictly increasing u, let u^{**} be a strictly increasing extension of v. Choosing $\epsilon > 0$ small enough, for $u = (1-\epsilon)u^* + \epsilon u^{**}$,

$$\alpha \sum_k u(x^{t,k}) + (1-\alpha) u(x^t) > \alpha \sum_k u(y^{t,k}) + (1-\alpha) u(\sum_{k=1}^{K_t} y^{t,k})$$

[37] The max exists since Q_Y^t is a finite set.

for all t and any $y^{t,1} \times \cdots \times y^{t,K_t} \in B^{t,1} \times \cdots \times B^{t,K_t} \setminus \left\{x^{t,1} \times \cdots \times x^{t,K_t}\right\}$, establishing that $\mathcal{D}$ is rationalized by α-PNB.[38]

Conditions under which $\mathcal{D}_Y^\alpha$ is rationalized by expected utility are well-known. Here, we follow Clark (1993). As standard, p is directly revealed preferred to q, written $p \succ q$, if $p = p^t$ and $q \in Q_Y^t \setminus \{p^t\}$ for some decision t. Define $p \succsim q$ if $p \succ q$ or $p = q$. We say that p is indirectly reveal preferred via independence to q, written $p \tilde{\succsim} q$, whenever there exist $r, p_1, \ldots, p_n, q_1, \ldots, q_n \in \Delta Y$, $\beta \in (0,1]$, and $\lambda_1, \ldots, \lambda_n > 0$ so that

$$\beta p + (1-\beta) r = \sum_{i=1}^{n} \lambda_i p_i$$

$$\beta q + (1-\beta) r = \sum_{i=1}^{n} \lambda_i q_i$$

$$p_i \succsim q_i \text{ for all } i = 1, \ldots, n$$

$$\sum_{i=1}^{n} \lambda_i = 1$$

and $p \tilde{\succ} q$ holds whenever at $p_i \succ q_i$ for some i. The Linear Axiom of Revealed Preference (LARP) is that $q \tilde{\succsim} p$ implies that $p \tilde{\succ} q$ does not hold. Theorem 3 of Clark (1993), combined with the finiteness of $\mathcal{D}^\alpha$, implies that $\tilde{\succsim}$ has an expected utility rationalization. Conclude $\mathcal{D}^\alpha$ satisfies LARP if and only if $\mathcal{D}$ is rationalized by α-partial-narrow bracketing.

The following algorithm inputs $\mathcal{D}^\alpha$ and outputs 1 if and only $\mathcal{D}$ is rationalized by α-partial-narrow bracketing for any rational number $\alpha \in [0,1]$. For concreteness, label $Y = \{y_1, y_2, \ldots, y_m\}$. We identify each $p \in \Delta Y$ with the $p \in \mathbb{R}^m$ so that $p_i = p(y_i)$.

Define the matrix A with a column equal to $p^t - q$ for each $q \in Q_Y^t \backslash p^t$ and every t. A is an $D \times m$ matrix interpreted as strict preference, where D is the number of data points extracted. Let $\vec{1}$ be a $D \times 1$ matrix of ones and

$$\hat{A} = \begin{bmatrix} A \\ \vec{1} \end{bmatrix}.$$

Let $\hat{b}$ be an $m+1$ vector with the first m components 0 and the last 1.

[38]This approach is similar to that of Polisson *et al.* (2020).

Note that if $p\tilde{\succsim}q$, then there exists $\beta > 0$,$\lambda_i \geq 0$, and $p_i \succsim q_i$ for $i = 1, ..., m$ so that

$$\beta(p - q) = \sum_{i=1}^{m} \lambda_i (p_i - q_i)$$

and $\sum_{i=1}^{m} \lambda_i = 1$. Each $p_i - q_i$ corresponds to a column of A, so this is equivalent to $p - q \in C = cone(co(\{A_1, \ldots, A_D\}))$ where A_i is the ith column of A. LARP holds when $p - q \in C$ implies $q - p \notin C$, or, equivalently, if and only if $0 \notin C$. That is, there does not exist $\phi^1, \ldots, \phi^D \geq 0$ so that $\sum \phi^i = 1$ so that $\sum_{i=1}^{D} \phi^i A_i = 0$, which is equivalent to $\hat{A}\phi = \hat{b}$.[39]

Solve the system

$$\min_{y \in \mathbb{R}^{m+1}} y^T \hat{b}$$
$$s.t. \hat{A}^T y \geq 0$$

for y. The Ellipsoid algorithm provides a polynomial-time solution to this. Let y^* be the solution, if one exists. There are two cases:

Case 1: There is no solution to $\hat{A}^T y \geq 0$ or y^* satisfies $y^{*T}\hat{b} \geq 0$. Then, Farkas's Lemma implies there exists $\phi \geq 0$ so that $\hat{A}\phi = \hat{b}$. LARP fails by the above, so return 0.

Case 2: The program is unbounded below or $y^{*T}\hat{b} < 0$. Then, Farkas's Lemma implies there **does not** exist any $\phi \geq 0$ so that $\hat{A}\phi = \hat{b}$. LARP is satisfied by the above, so return 1.

To specialize the above to PNB-PE, we perform the analysis part by part rather than decision by decision. Let $p^{t,k} = (\alpha, x^t; (1 - \alpha), x^{t,k})$, $Y \subseteq \mathbb{R}^n$ be a finite set that includes the union of the supports of the lotteries $p^{t,k}$, and

$$Q_Y^{t,k} = \left\{(\alpha, y; (1 - \alpha), y') \in \Delta Y : \exists x \in B^{t,k} \text{ so that } y' \leq x + x^{t,-k} \,\&\, y \leq x\right\}$$

Repeating the above with the ancillary dataset

$$\mathcal{D}_Y^{PE,\alpha} = \left\{\left(p^{t,k}, Q^{t,k}\right)\right\}_{(t,k)}$$

[39]To extend to allow for some weak preference, e.g. implications of symmetry, let B be an $D' \times m$ matrix with **columns** given by $p - q$ where p and q represent comparisons including weak preference and $\vec{0}$ be a $D' \times 1$ matrix of zeroes. Repeat the above with $\hat{A} = \begin{bmatrix} A & B \\ \vec{1} & \vec{0} \end{bmatrix}$.

replacing $\mathcal{D}^{\alpha}$ establishes the result.

□

Theorem 3 (Identification of α.)**.** *Suppose that u is a known differentiable and strictly quasi-concave function and $\mathcal{D}$ is rationalized by partial-narrow bracketing. Further suppose that for some t,k, $B^{t,k}$is a Walrasian budget set for prices $p^{t,k}$, and $n=2$, $\frac{p_1^{t,1}}{p_2^{t,1}} \neq \frac{p_1^{t,2}}{p_2^{t,2}}$, $\lim_{x_1 \to 0^+} \frac{\partial u(x_1,x_2)}{\partial x_1} = +\infty$ for all $x_2 > 0$, and $\lim_{x_2 \to 0^+} \frac{\partial u(x_1,x_2)}{\partial x_2} = +\infty$ for all $x_1 > 0$.*
Then, there exists a unique α that rationalizes choices.

Proof of Theorem 3. Suppose the assumptions of the Theorem are satisfied. In the partial-narrow bracketing model, the marginal utility per dollar of spending on good j in decision k is given by:

$$\frac{1}{p_j^{t,k}}\left((1-\alpha)\frac{\partial u(x^{t,1}+x^{t,2})}{\partial x_j} + \alpha\frac{\partial u(x^{t,k})}{\partial x_j}\right)$$

for $k=1,2$. By the limit conditions, if $\alpha > 0$, then $x_j^{t,k} > 0$ for $j=1,2$ and $k=1,2$; thus $x_1^{t,k}=0$ or $x_2^{t,k}=0$ implies $\alpha = 0$. If $\frac{1}{p_2^{t,k}}\partial u(x^{t,k})/\partial x_2 = \frac{1}{p_1^{t,k}}\partial u(x^{t,k})/\partial x_1$ for both k, then set $\alpha = 1$. Now suppose that $\frac{1}{p_2^{t,k}}\partial u(x^{t,k})/\partial x_2 \neq \frac{1}{p_1^{t,k}}\partial u(x^{t,k})/\partial x_1$ and an interior allocation is chosen in budget set k. The first-order condition for an interior maximizer equates the marginal utility per dollar across the two goods; rearranging this condition, we obtain

$$\alpha = \frac{\frac{1}{p_1^{t,k}}\partial u(x^{t,1}+x^{t,2})/\partial x_1 - \frac{1}{p_2^{t,k}}\partial u(x^{t,1}+x^{t,2})/\partial x_2}{\frac{1}{p_1^{t,k}}\partial u(x^{t,1}+x^{t,2})/\partial x_1 - \frac{1}{p_2^{t,k}}\partial u(x^{t,1}+x^{t,2})/\partial x_2 + \frac{1}{p_2^{t,k}}\partial u(x^{t,k})/\partial x_2 - \frac{1}{p_1^{t,k}}\partial u(x^{t,k})/\partial x_1}$$

for $k=1,2$. It remains to verify that the expression on the right-hand side always lies in the interval $(0,1)$. Since $\frac{1}{p_2^{t,k}}\partial u(x^{t,k})/\partial x_2 \neq \frac{1}{p_1^{t,k}}\partial u(x^{t,k})/\partial x_1$, the solution satisfies $(1-\alpha)\left(\frac{1}{p_1^{t,k}}\frac{\partial u(x^{t,1}+x^{t,2})}{\partial x_1} - \frac{1}{p_2^{t,k}}\frac{\partial u(x^{t,1}+x^{t,2})}{\partial x_2}\right) = \alpha\left(\frac{1}{p_2^{t,k}}\frac{\partial u(x^{t,k})}{\partial x_2} - \frac{1}{p_1^{t,k}}\frac{\partial u(x^{t,k})}{\partial x_1}\right)$. Thus, if $\alpha \in (0,1)$, the denominator and numerator have the same sign, and the denominator is strictly larger in absolute value – therefore the expression is well-defined and thus α is identified from choices.

□

Joint implications of narrow and broad bracketing. To study the joint implications of broad and narrow bracketing we consider endowment bracketing in a rich domain. We thus suppose that we observe an infinite dataset on all decisions that each consist of a binary part and a second degenerate part summarized by its only available bundle z. Suppose that when a person has bundle $z \in \mathbb{R}^n_+$ in their second part and must make a binary choice, they apply a complete, transitive, continuous, and strictly increasing binary relation $\succsim_z$ (call such a binary relation a *well-behaved preference*). In this domain, we say that a person *broadly brackets* if $x \succsim_z y \iff x + z \succsim_{\underline{0}} y + z$; a person *narrowly brackets* if $x \succsim_z y \iff x \succsim_{z'} y$ for all $z' \in \mathbb{R}^n_+$.

Theorem 4. *The family of well-behaved preferences $\{\succsim_z\}_{z\in\mathbb{R}^n_+}$ satisfies both narrow and broad bracketing if and only if there exists a vector $u \in \mathbb{R}^n_{++}$ such that $x \succsim_z y$ if and only if $x \cdot u \geq y \cdot u$.*

Proof. The "if" direction is immediate. So now consider the "only if" direction.We do so by constructing an additive utility representation for $\succsim_{\underline{0}}$. Since $\succsim_{\underline{0}}$ is a well-behaved preference relation, for each $x \in \mathbb{R}^n_+$ there exists a unique number $\kappa \in \mathbb{R}_+$ such that $x \sim_{\underline{0}} \kappa e$ where $e = (1, 1, \ldots, 1)$; define $u(x) = \kappa$ for each x and the corresponding κ. By strict monotonicity, u represents $\succsim_{\underline{0}}$, and by continuity, u is continuous. Pick any $x, y \in \mathbb{R}^n_+$. By NB, $x \sim_y u(x)e$ and $y \sim_{u(x)e} u(y)e$, so by BB and transitivity $x + y \sim_0 y + u(x)e \sim_0 u(x)e + u(y)e$, so $u(x+y) = u(x)+u(y)$. Since x, y were arbitrary, $u(x) = u \cdot x$ for some $u \in \mathbb{R}^n_{++}$ by Theorem 1 in Chapter 5 of Aczél (1966). □

Derivations of predictions and their experimental implementations. First, observe that if relation P is acyclic, it must be anti-symmetric.

Derivation of Prediction 1 (NB-WARP). Suppose $x^{t',k'} \in B^{t,k} \subseteq B^{t',k'}$. By the definition of P in NB-SARP, $x^{t',k'} P x$ for all $x \in B^{t',k'} \backslash \{x^{t',k'}\}$. Since $B^{t,k} \subseteq B^{t',k'}$, it follows that $x^{t',k'} P x$ for all $x \in B^{t,k} \backslash \{x^{t',k'}\}$. But if $x^{t,k} \neq x^{t',k'}$, the definition of P would imply $x^{t,k} P x^{t',k'}$, which would violate acyclicity. Thus $x^{t,k} = x^{t',k'}$.

Discretized implementation in our experiment. In all of our tests of NB-WARP, we have $B^{t,k} = B^{t',k'}$ exactly. We thus test NB-WARP exactly, without needing to adjust for discreteness.

Derivation of Prediction 2 (BB-WARP). Suppose $x^{t'} \in B^t \subseteq B^{t'}$. By the definition of P in BB-SARP, $x^{t',k'} P x$ for all $x \in B^{t'} \backslash \{x^{t'}\}$. Since $B^t \subseteq B^{t'}$, it follows that $x^{t'} P x$ for all $x \in B^t \backslash \{x^{t'}\}$. But if $x^t \neq x^{t'}$, the definition of P would imply $x^t P x^{t'}$, which would violate acyclicity. Thus $x^t = x^{t'}$.

Discretized implementation in our experiment. In the Risk and Social experiments, $\text{co}B^{1,1} + \text{co}B^{1,2} \subseteq \text{co}B^{2,1}$. In particular, $B^{2,1} = \{(\$0,\$28)\,;(\$2,\$26)\,;\ldots;(\$28,\$0)\}\}$, whereas $B^{1,1} + B^{1,2} = \{(\$0,\$28)\,;(\$1,\$27)\,;\ldots;(\$16,\$12)\,;(\$17,\$10.80)\,;\ldots;(\$26,\$0)\}$. Consider two cases.

If $x_A^{2,1} \leq \$16$, the bundle chosen in $B^{2,1}$ is exactly affordable in $B^{1,1} + B^{1,2}$. The caveat is that the set $B^{1,1} + B^{1,2}$ has a higher resolution in this range. We do not adjust for this, implicitly interpreting their choice from $B^{2,1}$ as revealing their preferences over the $\text{co}B^{2,1}$. However, if this leads a subject to fail BB-WARP, if their preferences are well-behaved, they would choose in $B^{2,1}$ one of the closest bundles to their preferred bundle in $\text{co}B^{2,1}$ and be within one error away from passing BB-WARP.

However, if $x_A^{2,1} > \$16$, then the person reveals a sufficiently strong preference for person/state A over B that their desired bundle in B^2 is not affordable in $\text{co}B^{1,1} + \text{co}B^{1,2}$ and thus they trivially pass BB-WARP.

Derivation of Prediction 3 (BB-Mon). Suppose $x_1^{t,2} > 0$, $x_2^{t,1} > 0$, and $p_1^{t,1} \leq p_2^{t,1}$, $p_1^{t,2} > p_2^{t,2}$. Let $\epsilon = \min\{x_1^{t,2}, x_2^{t,1}\}$. Consider the alternative pair of choices $(y^{t,1}, y^{t,2})$ given by $y^{t,1} = \left(x_1^{t,1} + \frac{p_2^{t,1}}{p_1^{t,1}}\epsilon, x_2^{t,1} - \epsilon\right)$ and $y^{t,2} = \left(x_1^{t,2} - \epsilon, x_2^{t,2} + \frac{p_1^{t,2}}{p_2^{t,2}}\epsilon\right)$. By construction, $(y^{t,1}, y^{t,2})$ is affordable. We have that

$$y_1^{t,1} + y_1^{t,2} = x_1^{t,1} + x_1^{t,2} + \frac{p_2^{t,1} - p_1^{t,1}}{p_1^{t,1}}\epsilon \geq x_1^{t,1} + x_1^{t,2}, \text{ and}$$

$$y_2^{t,1} + y_2^{t,2} = x_2^{t,1} + x_2^{t,2} + \frac{p_1^{t,2} - p_2^{t,2}}{p_2^{t,1}}\epsilon > x_2^{t,1} + x_2^{t,2}$$

where the inequalities respectively follow from $p_1^{t,1} \leq p_2^{t,1}$ and $p_1^{t,2} > p_2^{t,2}$. But $x^t P y^t$ would violate monotonicity; thus such an $x^{t,1}, x^{t,2}$ pair could not pass BB-SARP.

Discretized implementation in our experiment. The argument in the proof applies with minimal modification to our discretized experiment, and thus we apply the conditions directly.

Online Appendix:

Revealing Choice Bracketing

Andrew Ellis and David J. Freeman

Appendix B. Supplementary Data Visualizations

Risk and Social Experiments. First, we plot histograms of allocations in each part in Risk and Social Experiments (Figures 2 and 3). Notice that while the two experiments have a similar number of subjects (99 and 102 respectively), the y-axis scale differs across the two histograms. This reflects greater variation in behavior in the Risk than in the Social Experiment in D2 and D4. In both experiments, the vast majority of subjects make an equal allocation between assets/persons A and B, consistent with strict symmetric preferences. In Risk, 69 subjects make equal allocations in D4 (79 are within one token of doing so) while in Social the corresponding number is 90 subjects (97 within one token). While some variation in behavior is the norm in experiments, prior normative, behavioral, and experimental work on risk preferences generally suggests that risk aversion over 50-50 lotteries is not just strongly modal but a nearly universal. This illustrates the value of including D2, D4, and D5 in our Risk Experiment to enable us to apply WARP-style tests which do not require assumptions about preferences.

Figure 2. Allocations in Risk by part

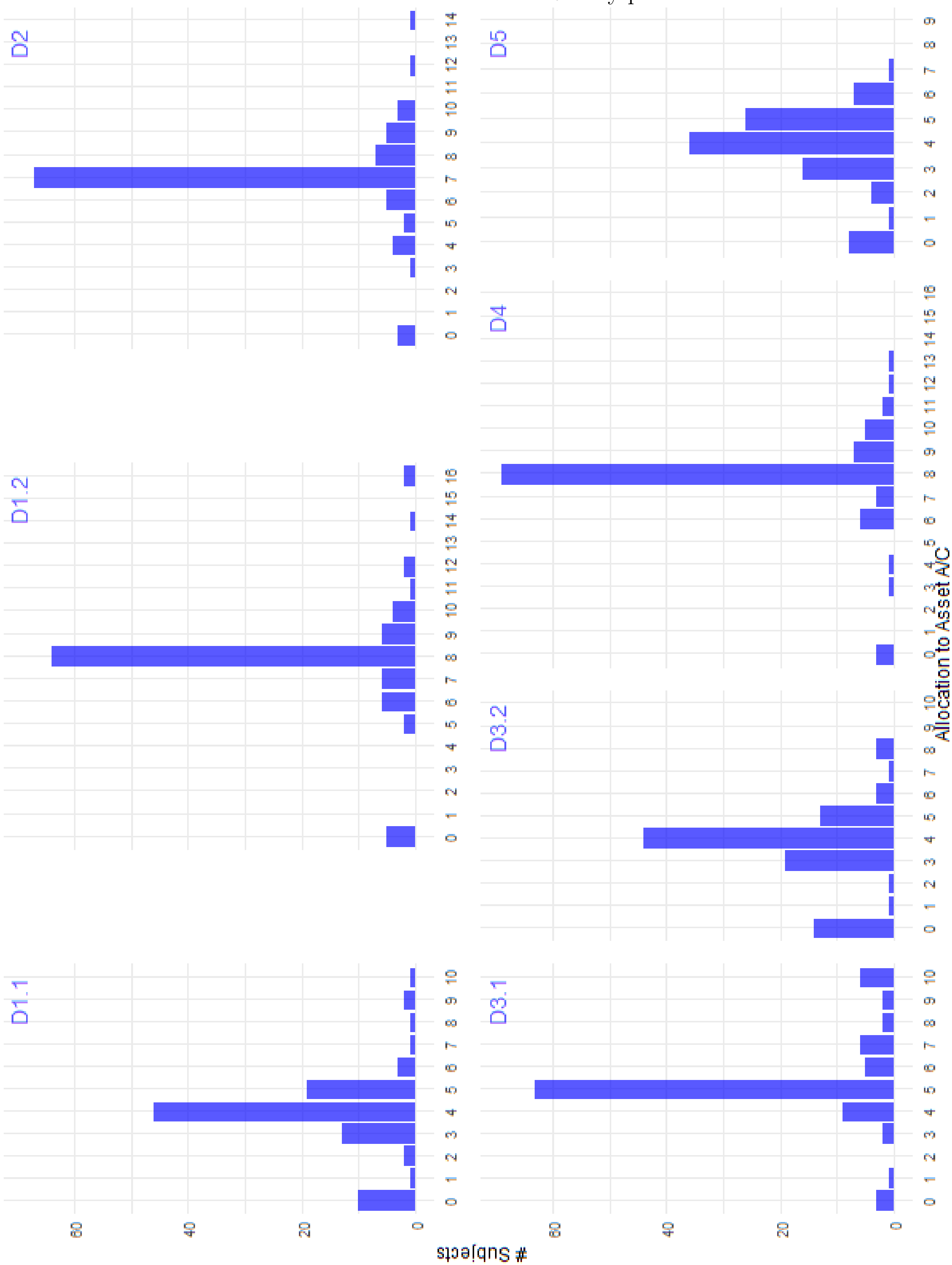


Figure 3. Allocations in Social by part

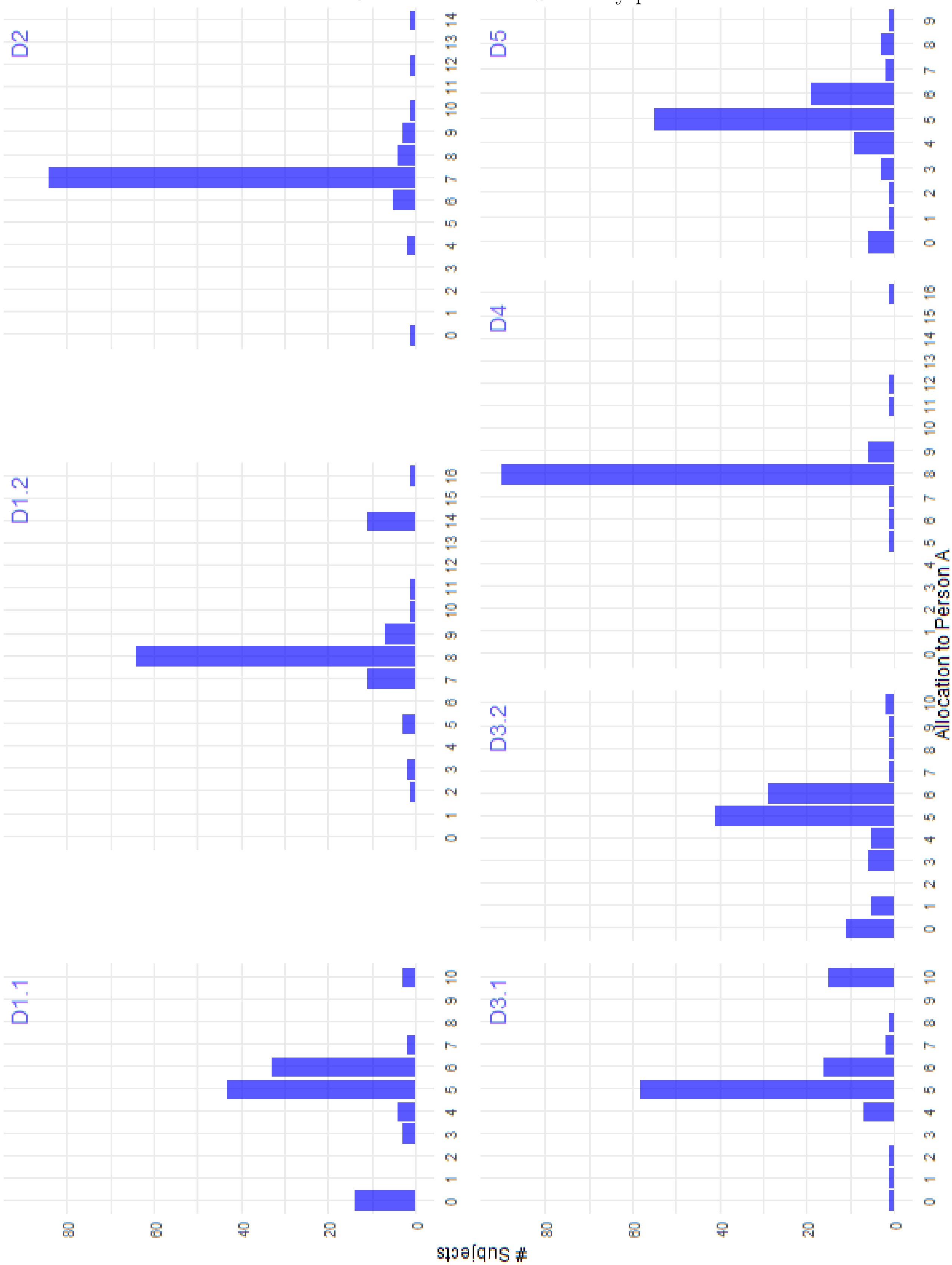


Figure 4 plots the aggregate (i.e. decision-level) allocations to die rolls 1-3/4-6 (Risk) and Person A/B (Social) in Decision 1 (shown in red) and in Decision 2 (shown in blue); the size of each circle is proportional to the number of subjects making an allocation. Our BB-WARP is based on a comparison of these allocations for each individual. While the blue dots (D2) lie on a line with slope -1 extending from (0, 28) to (28,0), almost all of the mass in red (D1) lies below that line – indicating subjects who made D1 choices that resulted in a dominated overall bundle. Since it is impossible in Decision 1 to obtain an aggregate bundle that allocates strictly more than \$16 to die roll 1-3/Person A, the very small number of subjects (10 in Risk, 6 in Social) who make such an allocation in Decision 2 automatically pass BB-WARP.

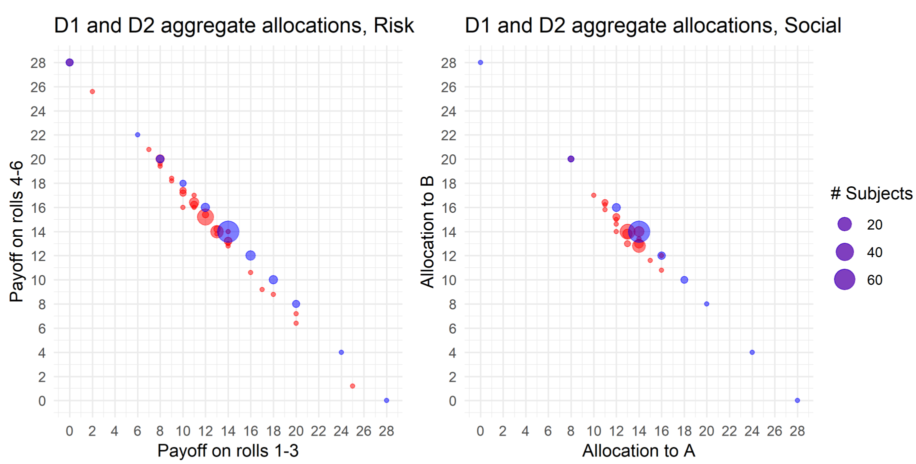


FIGURE 4. Aggregate allocations in D1 (red) and D2 (blue)

Figure 5 plots the individual-level allocations to Asset A/C in D1.1, D3.2, and D1.2 against their allocation in D5, D5, and D4 respectively. NB-WARP tests that the person makes the same allocation in each pair of parts, and is satisfied by all dots on line through (0,0) with a slope of 1.

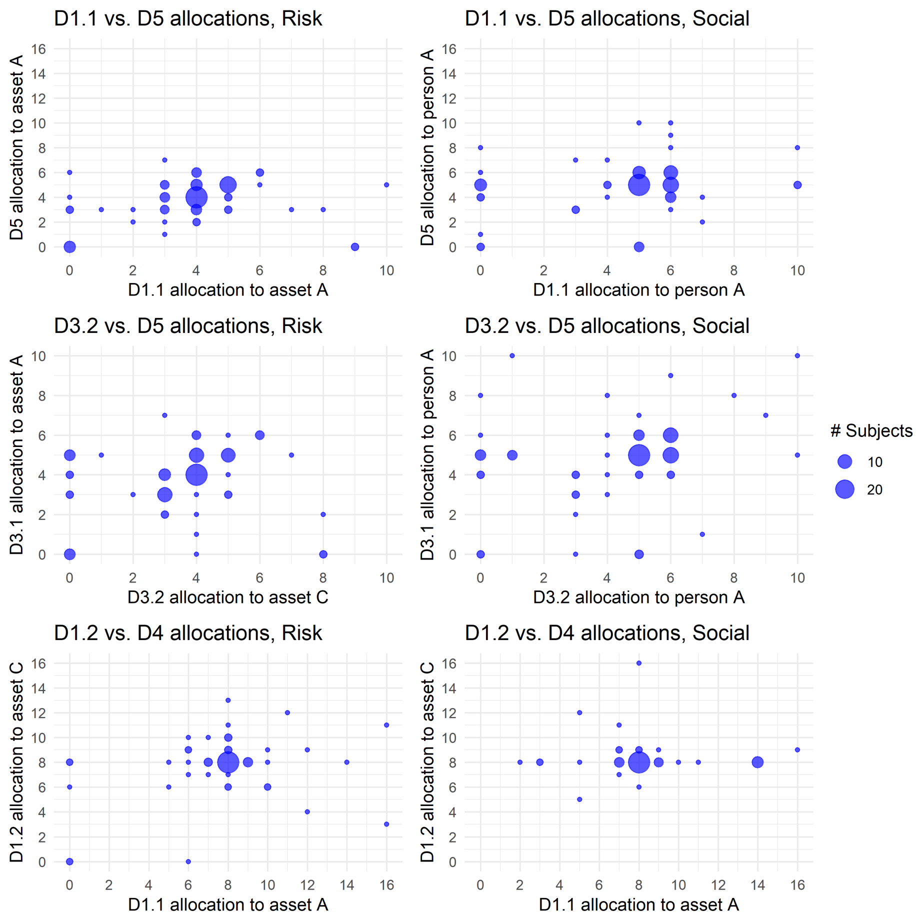
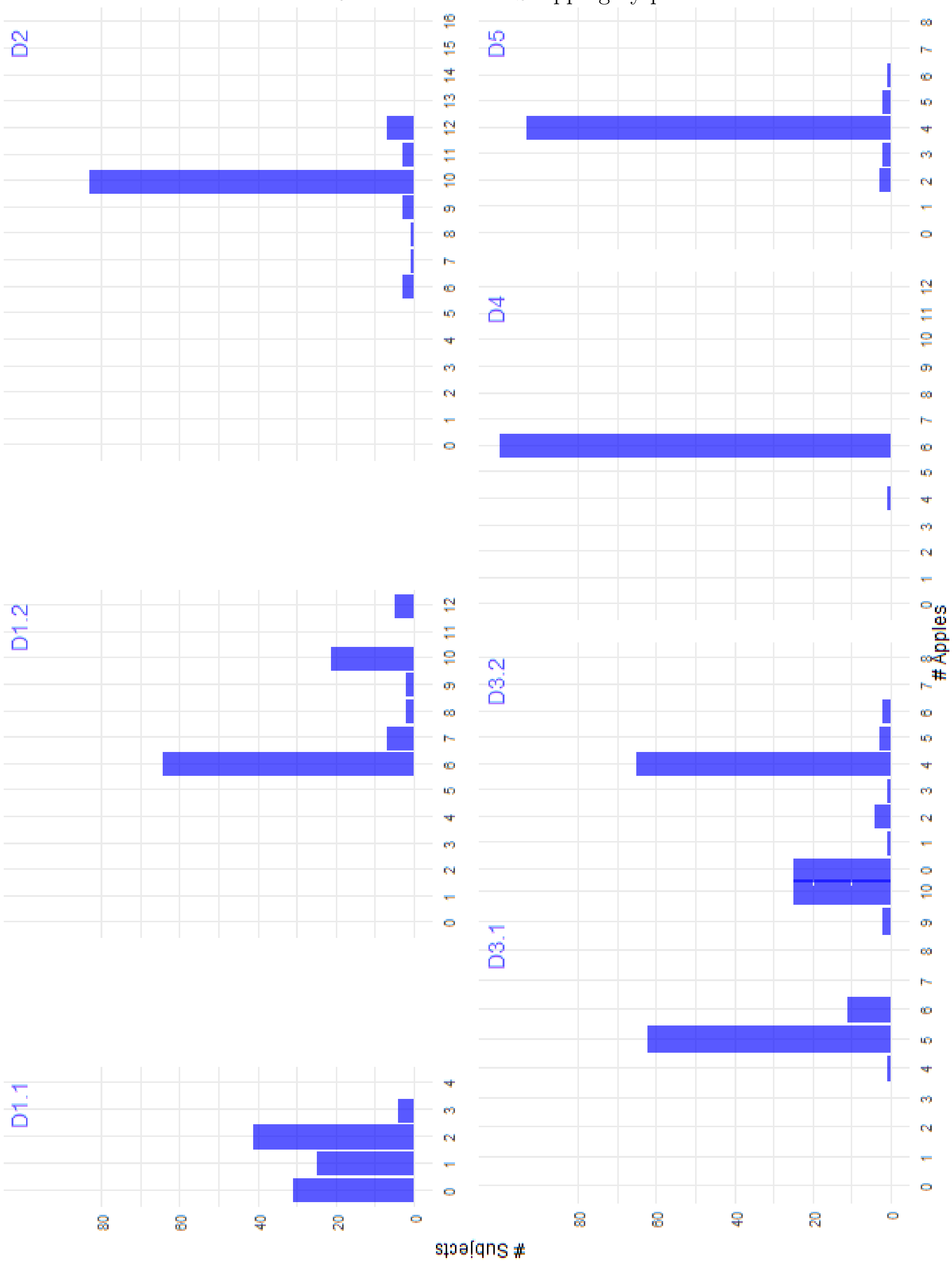


FIGURE 5. Comparison of identical parts

Shopping Experiment. Figure 6 plots histograms of purchases in each part. The histograms for D2, D4, and D5, which each have a single part, show very good adherence to value maximization: 91% of these allocations maximize payoffs exactly, and there are only 4 allocations (all in D2) in which subjects are more than two errors from the optimal allocation. The histograms for D1.1, D1.2, D3.1, and D3.2 illustrate close adherence to the predictions of full narrow bracketing with the exception that many subjects purchase two apples in D1.1 when narrow bracketing predicts they should buy only one,[40] and broad bracketing implies they should buy no apples.

[40]Note that since the price of oranges is half that of apples in D1.1 and lines correspond to numbers of oranges here, this counts as two errors relative to the narrow bracketing prediction. For a narrow bracketer, this is a very “small” mistake in terms of narrowly-assessed payoffs.

FIGURE 6. Allocations in Shopping by part



Further notes. With the exception of the output for the symmetric BB-SARP, NB-SARP, and PNB and PNBPE Algorithms, results in this paper were computed in R (R Core Team, 2020), and made use of the tidyr package (Wickham & Henry, 2020). Plots generated using ggplot2 (Wickham, 2016) and ggpubr (Kassambara, 2020).

Appendix C. Tests Assuming Symmetry

C.1. **Testable conditions.** We present the following testable implications of narrow, broad, and partial-narrow bracketing when underlying preferences are required to be symmetric.

Prediction 4 (NB-Sym)**.** Suppose $\mathcal{D}$ is rationalized by symmetric narrow bracketing and each $B^{t,k}$ is a Walrasian budget set for prices $p^{t,k}$.
If $p_j^{t,k} \geq p_i^{t,k}$, then $x_i^{t,k} \geq x_j^{t,k}$.

With narrow bracketing, symmetry's implications are straightforward. The subject should purchase at least as much of the cheaper good. With broad and partial-narrow bracketing, the implications are more subtle.

Prediction 5 (BB-Sym)**.** Suppose $\mathcal{D}$ is rationalized by symmetric broad bracketing, and each $B^{t,k}$ is a Walrasian budget set for prices $p^{t,k}$ and income $I^{t,k}$. Then,
(i) if $B^t = B^{t,1} + B^{t,2}$,$p_1^{t,1} = p_2^{t,1}$, and $p_i^{t,2} > p_j^{t,2}$, then $x_i^{t,1} + x_i^{t,2} \leq x_j^{t,1} + x_j^{t,2}$; if in addition $\frac{I^{t,2}}{p_j^{t,2}} \leq \frac{I^{t,1}}{p_i^{t,1}}$, then $x_i^{t,2} = 0$, and
(ii) if $B^t = B^{t,1}$ and $p_i^{t,1} \geq p_j^{t,1}$, then $x_j^{t,1} \geq x_i^{t,1}$.

Prediction 6 (PNB-Sym)**.** Suppose $\mathcal{D}$ is rationalized by symmetric partial-narrow bracketing and each $B^{t,k}$ is a Walrasian budget set for prices $p^{t,k}$. Then,
(i) if $B^t = B^{t,1} + B^{t,2}$, $p_1^{t,1} = p_2^{t,1}$ and $p_i^{t,2} > p_j^{t,2}$, then $x_j^{t,2} \geq x_i^{t,2}$ and $x_j^t \geq x_i^t$, and
(ii) if $B^t = B^{t,1}$ and $p_1^{t,1} \geq p_2^{t,1}$, then $x_2^{t,1} \geq x_1^{t,1}$.

For an intuition, recall the logic of comparative advantage. The opportunity cost of consuming good i from $B^{t,1}$ is lower than in $B^{t,2}$, and vice versa for consuming good j. A DM should purchase at least as much of the good in the part where it is cheapest relative to the other good. With broad bracketing, this specialization is extreme: if a subject purchases positive amounts of good i from the second part, then they must exhaust the budget of the first on good i. Otherwise, they forgo the opportunity to consume more of each. The DM purchases the good only where it is cheapest, until switching to purchasing the other good. With partial-narrow, the specialization is less extreme as the DM weighs the effect of less consumption of the more expensive good in the part as well as in the decision overall.

C.2. **Tests of NB-, BB-, and PNB-Sym.** Symmetric preferences are natural in our experimental setup, and in the decisions with equal prices, such as Decision 2, the majority of subjects allocate evenly. By testing this property's implications in other choices, we obtain more powerful tests that do not rely on cross-decision comparisons. This power comes at the cost of having to jointly test bracketing and that underlying preferences are symmetric. To the extent that these properties are compelling in our environment, the next set of tests distinguish the three models of bracketing.

	Risk			Social		
# errors	0	1	2	0	1	2
NB-Sym (D1)	62	73	84	41	72	82
NB-Sym (D3)	60	74	82	34	76	82
NB-Sym (both)	46	57	65	22	46	67
BB-Sym (D1)	1	1	1	11	11	12
BB-Sym (D3)	6	7	8	14	14	14
BB-Sym (both)	0	0	0	10	10	10
PNB-Sym (D1)	90	92	93	63	95	99
PNB-Sym (D3)	91	94	96	67	96	97
PNB-Sym (both)	86	88	92	46	77	93
# subjects	99			102		

Entries count the # of subjects who pass each test at the listed error allowance.

Table 8. Tests of NB-, BB-, and PNB-Sym

We thus test NB-Sym, BB-Sym, and PNB-Sym, using all of each condition's implications for a given decision. In Decision 1 tests, we find that 63% and 40% of subjects pass NB-Sym in the Risk and Social Experiments respectively. These pass rates increase to 74% and 71% if we allow for one error. We find quantitatively similar results for Decision 3. Pass rates go down to 46% and 22% when we require a subject to pass NB-Sym in both Decisions 1 and 3 – but these numbers increase to 58% and 45% when we allow for one error across both decisions. Thus most, but not all, subjects pass this test of the conjunction of narrow bracketing with symmetric preferences.

Turning to tests of broad bracketing, 1% and 11% of subjects pass BB-Sym in Decision 1 of the Risk and Social Experiments respectively. Allowing for one error does not increase pass rates at all, though one additional subject is within two errors of passing. We obtain slightly higher pass rates (6% and 14%) in Decision 3. This discrepancy can be traced to the relative strength of BB-Sym in these decisions. It makes a point prediction in Decision 1 but allows two possible choices in Decision 3. No subjects pass BB-Sym in both Decisions 1 and 3 for Risk, even when allowing for two errors. In Social, 10% of subjects pass, which does not change when allowing for up to two errors. Thus these tests suggest that only a small minority of subjects are close to consistent with broad bracketing.

Our PNB-Sym tests of partial-narrow bracketing have higher pass rates, as expected – 91% and 62% for Risk and Social Experiments respectively in Decision 1. Allowing one error raises pass rates to include the vast majority of subjects – 93% and 93% respectively. We obtain quantitatively similar results for Decision 3. However, allowing one error, only 18% and 12% pass PNB-Sym and neither BB-Sym nor NB-Sym in Decision 1 in the two experiments. We find that 87% and 45% pass PNB-Sym in both decisions, while 89% and 75% do so when allowing for one error. However, this means that, when allowing for one error of tolerance, only 31% and 20% of subjects pass PNB-Sym in both Decision 1 and 3 who pass neither BB-Sym nor NB-Sym. Thus, partial-narrow bracketing only somewhat helps to account for behavior in our experiments, in spite of the model's relatively weak implications in our experiment without parametric assumptions about utility.

Result 4. *When allowing for one error in the Risk and Social Experiments, 58% and 45% of subjects pass NB-Sym and 0% and 10% pass BB-Sym; only 31% and 20% pass PNB-Sym but not NB- nor BB-Sym.*

Appendix D. Power

For each test presented in Tables 3, 4, and 5 we compute the probability that randomly-generated choices would pass each. This approach to analyzing the power of revealed preference tests follows Bronars (1987), and has been used in Andreoni & Miller (2002, p. 744) and Choi *et al.* (2007, p. 1927), among other papers.

	Risk/Social		
# errors	0	1	2
NB-WARP (D1.1 and D5)	0.091	0.256	0.405
NB-WARP (D1.2 and D4))	0.059	0.170	0.273
NB-WARP (D3.2 and D5)	0.091	0.256	0.405
NB-WARP (D1.1 and D3.2)	0.091	0.256	0.405
NB-WARP (all)	0.0005	0.004	0.014
BB-WARP (D1 and D2)	0.427	0.517	0.591
BB-Mon (D1)	0.144	0.278	0.401
BB-Mon (D3)	0.174	0.331	0.471
BB-Mon (both)	0.025	0.071	0.134

Table 9. Probability of Random Choice Passing NB-/BB- WARP tests

	Risk/Social		
# errors	0	1	2
NB-SARP	9×10^{-8}	1×10^{-6}	1×10^{-5}
BB-SARP	2×10^{-7}	2×10^{-6}	9×10^{-6}
PNB	6×10^{-4}	3×10^{-3}	9×10^{-3}
PNB-PE	2×10^{-3}	1×10^{-2}	3×10^{-2}

Table 10. Probability of Random Choice Passing Full Tests

	D1			D3		
# errors	0	1	2	0	1	2
NB	0.009	0.043	0.111	0.007	0.035	0.091
BB	0.009	0.034	0.077	0.007	0.021	0.042
PNB	0.043	0.179	0.333	0.035	0.126	0.224
PNB-PE	0.043	0.179	0.333	0.041	0.147	0.238

	Both			Full		
# errors	0	1	2	0	1	2
NB	$6{\times}10^{-5}$	5×10^{-4}	0.002	2×10^{-8}	3×10^{-7}	2×10^{-6}
BB	$6{\times}10^{-5}$	4×10^{-4}	0.001	2×10^{-8}	2×10^{-7}	2×10^{-6}
PNB	5×10^{-4}	0.003	0.011	2×10^{-7}	2×10^{-6}	1×10^{-5}
PNB-PE	6×10^{-4}	0.004	0.012	2×10^{-7}	3×10^{-6}	2×10^{-5}

TABLE 11. Probability of Random Choice Passing each Shopping Test

APPENDIX E. EXTREMENESS AVERSION

In our Social/Risk experiments, BB-Mon requires extreme allocations in D1 and D3. In D1 (D3), it requires that either that all 10 tokens are allocated to person/asset B (B/D) in D1.1 (D3.2) when it offers a better return for the same price, or that no tokens are allocated to person/asset B (D) in D1.2 (D3.1).[41] In the main text, we say that a subject is consistent with *Extremeness-Averse (EA-)BB-Mon* if they are within 2 tokens of making an extreme allocation in those both of those two decisions. We picked 2 tokens because about a third of possible allocations are consistent with 3 tokens of slack but no errors, while more than half the choice space (12,055 out of 22,627 possible allocations) are consistent with 3 tokens of "slack" when allowing for 2 errors.

In our pen-and-paper shopping experiment, D3 requires two corner choices, while D1 only requires one corner choice. If extremeness aversion drove our results, we would expect more people consistent with broad bracketing in D1 than in D3. However, the rates of broad bracketing are similar across the two decisions, with between 20-27 subjects in D1 and either 23 or 24 subjects in D3 consistent with broad bracketing. The difference is not statistically

[41]Narrow bracketing predicts that at least 5 should go to asset/person B. All other subjects were at least as close to the narrow bracketing predictions as to broad bracketing predictions.

Table 12. Percent of all allocations consistent with EA-BB-Mon

	0 errors	1 errors	2 errors	3 errors	4 errors
0 slack	2.5	7.1	13.4	20.9	29.4
1 slack	9.2	17.2	26.1	35.5	45.2
2 slack	18.9	29.2	39.6	50.0	60.0
3 slack	30.5	42.1	53.3	63.8	73.3
4 slack	43.2	55.2	66.2	76.1	84.5
5 slack	56.0	67.6	77.8	86.4	93.1

Each cell indicates percentage of all possible allocations consistent with EA-BB-Mon in both D1 and D3 at indicated level of slack and error allowance in Social and Risk experiments.

Table 13. EA-BB-Mon in Risk

	BB-Mon			EA-BB-Mon			
	0 errors	1 errors	2 errors	0 errors	1 errors	2 errors	n
all	7	8	10	10	15	24	99
0 errors NB-SARP	1	1	1	1	1	2	23
1 errors NB-SARP	1	1	1	1	1	3	34
2 errors NB-SARP	1	1	1	1	1	4	43
Classified NB	3	3	3	3	5	9	77

Each cell indicates number of subjects within the subgroup indicated by first column that pass the test in first row with number of errors in second row.

Table 14. EA-BB-Mon in Social

	BB-Mon			EA-BB-Mon			
	0 errors	1 errors	2 errors	0 errors	1 errors	2 errors	n
all	12	12	12	12	13	15	102
0 errors NB-SARP	1	1	1	1	1	2	15
1 errors NB-SARP	1	1	1	1	1	2	36
2 errors NB-SARP	1	1	1	1	1	2	44
Classified NB	2	2	2	2	2	3	77

Each cell indicates number of subjects within the subgroup indicated by first column that pass the test in first row with number of errors in second row.

significant when allowing for either 0, 1, or 2 errors (all p-values greater than 0.7, Fisher's exact test). Again, this suggests that extremeness aversion does not affect the fraction of broad bracketers by too much.

Appendix F. Order effects

We varied the order of both the overall decisions and of the parts within each two-part decision (see Table 15). Subjects in the lab sessions either completed decisions in order D1-D5 or in order D4, D3, D2, D1, D5.[42] We perform every pairwise test of the effect of the order in which decisions were faced separately for each experiment, comparing sessions who completed that decision earlier to those who completed it later (Table 16). Every such rank-sum test yields a p-value exceeding 0.10 with the exceptions of D1.2 and D2 in the Shopping Experiment ($p = 0.07, 0.02$ respectively). We performed 21 such tests, did not hypothesize the existence of order effects, and standard corrections (like the Bonferroni correction) for multiple hypothesis testing would render these tests insignificant. We thus attribute these differences to sampling variation, though the former difference may reflect learning.

We also varied the order of parts within each of the multi-part decisions, D1 and D3, and performed rank-sum tests of whether order affects allocations, separately for each part. Two of these 12 tests return a p-value < 0.05, the D1.1 test for Risk ($p = 0.02$) and the D3.2 test for Social ($p = 0.05$).[43] We did not hypothesize any order effects and standard corrections for multiple testing would render these insignificant. Again, this is probably due to sampling variation.[44]

[42]There were 6 subjects in Social who completed a different order due to a printing error, but we saw no reason to exclude these subjects from the analysis. The order was fully randomized in online sessions.

[43]One might conjecture that some subjects first narrowly bracket their choice in the the first part they face, and then select a broadly-bracketed best-reply to their previous choice in the second part they face. This is most cleanly tested using order variation in our Shopping Experiment, and we illustrate with reference to D3. In D3, a subject who faced Part 1 first would buy 5 of each fruit in Part 1 and then best-reply to that in Part 2 by buying 2 apples and 9 oranges – only 1 of 50 subjects does this. A subject who faced Part 2 first would buy 4 apples and 6 oranges in it, then reply with 6 apples and 4 oranges in Part 1 – only 6 of 51 subjects do this. This suggests that the heuristic of narrow bracketing in the first part faced, then best-replying in the second part faced, cannot explain many subjects' decisions in our setting.

[44]The median allocations are the same for both orders of the parts for each of D1.1, D1.2, D3.1, and D3.2 in both Risk and Social. Mean allocations in Risk to Asset A in D1.1 are 3.5 when Part 1 is first vs. 4.3 when Part 2 is first. But in Social, the mean allocation to A in D1.1 is 4.7 for both order variations. And in Risk, we see an effect in the opposite direction in D3.2: the mean allocation of is 3.7 to Asset C when Part 2 is first versus 3.3 when Part 1 is first. In contrast, in D3.2 of Social, the mean allocation is 4.1 to A when Part 2 is first versus 5.0 when Part 2 is second. The lack of consistency of these effects both within each experiment and when comparing a part in Risk to the same part Social suggests that these effects do not merit a systematic explanation.

Table 15. Orders

Order		Risk	Social	Shopping
(D1.1,D1.2), D2, (D3.1,D3.2), D4, D5	1M	14	24	15
(D1.2,D1.1), D2, (D3.1,D3.2), D4, D5	1L	16	14	15
(D1.1,D1.2), D2, (D3.2,D3.1), D4, D5	1H	19	15	17
(D3.2,D3.1), D2, (D1.1,D1.2), D4, D5	2X[45]	0	6	0
D4, (D3.2,D3.1), D2, (D1.2,D1.1), D5	2F	15	14	16
D4, (D3.1,D3.2), D2, (D1.2,D1.1), D5	2L	18	16	20
D4, (D3.2,D3.1), D2, (D1.1,D1.2), D5	2H	17	13	18
Total		99	102	101

Table 16. Tests of Order Effects

Order Comparison		Risk, medians			Social, medians			Shopping, medians		
		O1	O2	p-value	O1	O2	p-value	O1	O2	p-value
D1.1, Orders O1 vs O2		4	4	0.37	5	5	0.14	1	1	0.32
D1.2, Orders O1 vs O2		8	8	0.21	8	8	0.44	6	6	0.07
D2, Orders O1 vs O2		7	7	0.69	7	7	0.62	10	10	0.02
D3.1, Orders O1 vs O2		5	5	0.76	5	5	0.98	5	5	0.23
D3.2, Orders O1 vs O2		4	4	0.17	5	5	0.95	4	4	0.17
D4, Orders O1 vs O2		8	8	1.00	8	8	0.30	6	6	0.36
D5, Orders O1 vs O2		4	4	0.17	5	5	0.65	4	4	0.49
D1.1, order of parts		4	4	0.02	5	5	0.71	2	1	0.23
D1.2, order of parts		8	8	0.55	8	8	0.43	6	6	0.81
D3.1, order of parts		5	5	0.33	5	5	0.59	5	5	0.68
D3.2, order of parts		4	4	0.27	5	5	0.05	4	4	0.71

Entries for p-values are for a rank-sum test of the null hypothesis of the same distribution of allocations in the two orders. The first 7 tests compared 1L, 1H, and 1M allocations to 2L, 2H, 2F, 2X allocations. The D1.1, order of parts test compares allocations in D1.1 in orders in which D1.1 comes before D1.2 to orders in which D1.2 comes first (i.e. 1M, 1H, 2H, 2X vs. 1L, 2L, 2F); remaining order of parts tests are analogous.

One might hypothesize that some subjects learn make decisions that are more consistent with broad bracketing in the second two-budget-set round they face. We test this more directly using Shopping Experiment data, by comparing the number of errors a subject makes in each of D1 and D3 relative to optimal decisions implied by broad bracketing with the induced value function. 51 subjects deviate less severely from broad bracketing in the second two-part round than in the first, while 36 exhibit the opposite pattern ($p = 0.13$,

[45]Note that order 2X was unintended: it was printed, and run, due to a copy-and-paste error. However, we saw no reason to exclude it from our analysis.

sign test); the average difference is 1.20 fewer errors in the second two-part round ($p = 0.01$, paired t-test). This suggests that such learning effects are relatively small.

Appendix G. Relation to Preregistration Plan

We preregistered analysis plans for the Risk, Social, and Shopping Experiments online (https://osf.io/5wzrg, https://osf.io/362py, https://osf.io/8mraq)

Compared to our preregistrations for Risk and Social, we proposed "risk/inequity averse" preferences, but we instead used symmetric. In our setting, risk/inequity aversion implies symmetry but not vice versa. Our NB-SARP and BB-SARP tests are the direct tests mentioned; both assume strict symmetric preferences. Where we mentioned "GARP" in the preregistration plan for partial-narrow bracketing, we should have said "SARP." This should be clear since our experiment is designed to have power to perform SARP but not GARP tests.

In our classification, we did not include expected value/risk-seeking/linear preferences as a separate category. We note that choices consistent with such preferences are rare, and linear preferences have low predictive power. Thus including this as a separate category would not substantially affect our classification.

Our preregistration for Shopping indicated a plan to apply similar WARP-style test as for Risk and Social. After reflection, we decided that since we had induced the payoff function, it made more conceptual sense to apply the direct tests that use that information rather than WARP/GARP tests. The approach in the paper allows sharper conclusions,

Appendix H. Experimental Materials

We provide instructions and quizzes for all experiments, sample decision sheets for each, and the payoff table for the Shopping Experiment. An experimental round of choices was stapled together with the cover sheet on the first page.

FIGURE 7. Risk: Instructions

Investment task

There will be five rounds of the investment task. The first page of each round will announce the number of accounts in that round. At the end of each round, raise your hand so that the experimenter can collect your decisions and give you the decision sheet for the next round. At the end of all rounds, one round will be randomly selected to be the "round that counts". You will be paid your earnings from the round that counts based on (and only based on) your decisions in that round. Since any round could be the round that counts, you should behave in each round as if it is the round that counts.

In each round of this task, you will buy risky investments in up to two different "investment accounts". Each investment generates a return that depends on a roll of a six-sided dice. You have a separate budget for each account that can be spent only in that account. The dice will be rolled once, and you receive the returns from all your investments in all accounts in that round.

Example

As an example, suppose that in the round-that-counts you have two accounts.

You have 20 ECU in Account 1, which has two investments available; each investment costs 1 ECU per unit.

One unit in Asset A pays

$0.40 if the dice roll is 1, 2, or 3;

$0.10 if the dice roll is 4, 5, or 6.

One unit in Asset B pays

$0.25 if the dice roll is 1, 2, or 3;

$0.25 if the dice roll is 4, 5, or 6.

You have 15 ECU in Account 2, which has two investments available; each investment costs 1 ECU per unit.

One unit in Asset C pays

$0.60 if the dice roll is 1 or 2;

$0.00 if the dice roll is 3, 4, 5, or 6.

One unit in Asset D pays

$0.30 if the dice roll is 1 or 2;

$0.30 if the dice roll is 3, 4, 5, or 6.

Suppose that

In Account 1: you allocate 5 ECU to Asset A and 15 ECU to Asset B;

In Account 2: you allocate 8 ECU to Asset C and 7 ECU to Asset D.

Then, if the dice roll is **2**, you will be paid:

$$5 \times \$0.40 + 15 \times \$0.25 + 8 \times \$0.60 + 7 \times \$0.30 = \$12.65.$$

Figure 8. Risk: Quiz

Please answer the following questions and raise your hand after you have done so.

Question.

Suppose that a round has two accounts. Do your purchases in Account 1 affect what items you can afford to purchase in Account 2?

YES / NO (highlight one)

Question.

Suppose that in a round of the experiment has two accounts. Account 1 has two assets available, A and B. Account 2 has two different assets available, C and D.

Each unit of Asset A pays $0.50 if the dice roll is 1 or 2 and $1.00 if the dice roll is 3, 4, 5, or 6;

Each unit of Asset B pays $1.00 if the dice roll is 1 or 2 and $0.50 if the dice roll is 3, 4, 5, or 6.

Each unit of Asset C pays $0.50 if the dice roll is 1 or 2 and $0.00 if the dice roll is 3, 4, 5, or 6;

Each unit of Asset D pays $0.00 if the dice roll 1 or 2 and $1.00 if the dice roll is 3, 4, 5, or 6.

Suppose that you invest as follows:

in Account 1, you invest 2 ECU in Asset A and 6 ECU in Asset B;

in Account 2, you invest 4 ECU in Asset C and 2 ECU in Asset D.

1. If this round determines your payment, then how much will you earn if the dice roll is 2?

   ____

2. If this round determines your payment, then how much will you earn if the dice roll is 6?

   ____

FIGURE 9. Risk: Cover Sheet for Round 1 (order 1M)

Subject # Session

Round 1

In round 1, you have 2 investment accounts.

Figure 10. Risk: D1.1 Decision Sheet

Investment Account 1

You have **10 ECU** available in Account 1. Two assets are available for purchase, Asset A and Asset B.

The price of Asset A is **1 ECU per unit**.

The price of Asset B is **1 ECU per unit**.

One unit in Asset A pays

$1.00 if the dice roll is 1, 2, or 3;

$0.00 if the dice roll is 4, 5, or 6.

One unit in Asset B pays

$0.00 if the dice roll is 1, 2, or 3;

$1.20 if the dice roll is 4, 5, or 6.

Please highlight a feasible combination of purchases of Asset A and Asset B from the list below.

0 units of Asset A and 10 units of Asset B.

1 unit of Asset A and 9 units of Asset B.

2 units of Asset A and 8 units of Asset B.

3 units of Asset A and 7 units of Asset B.

4 units of Asset A and 6 units of Asset B.

5 units of Asset A and 5 units of Asset B.

6 units of Asset A and 4 units of Asset B.

7 units of Asset A and 3 units of Asset B.

8 units of Asset A and 2 units of Asset B.

9 units of Asset A and 1 unit of Asset B.

10 units of Asset A and 0 units of Asset B.

Figure 11. Risk: D1.1 Decision Sheet

Investment Account 2

You have **16 ECU** available in Account 2. Two assets are available for purchase, Asset C and Asset D.

The price of Asset C is **1 ECU per unit**.

The price of Asset D is **1 ECU per unit**.

One unit in Asset C pays

$1.00 if the dice roll is 1, 2, or 3;

$0.00 if the dice roll is 4, 5, or 6.

One unit in Asset D pays

$0.00 if the dice roll is 1, 2, or 3;

$1.00 if the dice roll is 4, 5, or 6.

Please highlight a feasible combination of purchases of Asset C and Asset D from the list below.

0 units of Asset C and 16 units of Asset D.

1 unit of Asset C and 15 units of Asset D.

2 units of Asset C and 14 units of Asset D.

3 units of Asset C and 13 units of Asset D.

4 units of Asset C and 12 units of Asset D.

5 units of Asset C and 11 units of Asset D.

6 units of Asset C and 10 units of Asset D.

7 units of Asset C and 9 units of Asset D.

8 units of Asset C and 8 units of Asset D.

9 units of Asset C and 7 units of Asset D.

10 units of Asset C and 6 units of Asset D.

11 units of Asset C and 5 units of Asset D.

12 units of Asset C and 4 units of Asset D.

13 units of Asset C and 3 units of Asset D.

14 units of Asset C and 2 units of Asset D.

15 units of Asset C and 1 unit of Asset D.

16 units of Asset C and 0 units of Asset D.

Figure 12. Social: Instructions

Division task

There will be five rounds of a task where you will asked to allocate tokens between two other participants who will herein be labelled "person A" and "person B". They will not be told your identity, and you will not be told their identities. That is, you will remain completely anonymous to each other.

In each round of this task, you will have tokens in up to two different accounts. You decide how to allocate tokens between person A and person B in each account. The value per token allocated to each of A and B may vary across rounds and across accounts. You have a separate budget of tokens for each account that can be allocated only in that account. Payments for a given round will be determined by the sum of the value of all tokens allocated in all accounts in that round.

The first page of each round will announce the number of accounts in that round. At the end of each round, raise your hand so that the experimenter can collect your decisions and give you the decision sheet for the next round.

You and every other participant has numbered a sealed envelope at the beginning of the experiment. Each participant has been randomly allocated to a group and role (A or B); this is recorded in the envelope. The round that counts to determine your payment has also been randomly selected and recorded in each envelope. Your group has been randomly and anonymously matched to determine the payment of another group and one round of your choices will determine the earnings of person A and person B in that group. Since each round could be the round that counts and actually determines a two other subjects' payments, you should treat each round as if it is the round that counts.

Example

As an example, suppose that in the round-that-counts there are two accounts.

There are 10 tokens in Account 1.

One token is worth \$0.80 to A and \$0.60 to B.

There are 12 tokens in Account 2.

One token pays \$1.00 to A and \$0.20 to B.

Suppose that

In Account 1: you allocate 4 tokens to A and 6 tokens to B.

In Account 2: you allocate 2 tokens to A and 10 tokens to B.

Then,

A's earnings are $4 \times \$0.80 + 2 \times \$1.00 = \$5.20$;

B's earnings are $6 \times \$0.60 + 10 \times \$0.20 = \$5.60$.

Figure 13. Social: Quiz

Please answer the following questions and raise your hand after you have done so.

Question.

Suppose that a round has two accounts. Does your allocation in Account 1 affect what you have available to allocate in Account 2?

YES / NO (highlight one)

Question.

Suppose that a round of the experiment has two accounts.

In Account 1, each token pays $0.40 to A and $0.60 to B.

In Account 2, each token pays $0.30 to A and $0.40 to B.

Suppose that you invest as follows:

in Account 1, you allocate 2 tokens to A and 4 tokens to B;

in Account 2, you allocate 6 tokens to A and 1 token to B.

1. If this is the round that counts for this group, then how much will person A receive?

   ____

2. If this is the round that counts for this group, then how much will person B receive?

   ____

Figure 14. Social: D5 Decision Sheet

<u>Account 1</u>

You have **10 tokens** available in Account 1.

Each token allocated to A is worth $1.00.

Each token allocated to B is worth $1.20.

Please highlight a feasible allocation of tokens between A and B.

0 tokens for A and 10 tokens for B.

1 token for A and 9 tokens for B.

2 tokens for A and 8 tokens for B.

3 tokens for A and 7 tokens for B.

4 tokens for A and 6 tokens for B.

5 tokens for A and 5 tokens for B.

6 tokens for A and 4 tokens for B.

7 tokens for A and 3 tokens for B.

8 tokens for A and 2 tokens for B.

9 tokens for A and 1 token for B.

10 tokens for A and 0 tokens for B.

FIGURE 15. Shopping: Instructions

Shopping Task

There will be five rounds of the shopping task. At the end of all rounds of the experiment, one round will be randomly selected to be the "round that counts". You will be paid your earnings from the round that counts based on (and only based on) your decisions in that round. Since any round could be the round that counts, you should behave in each round as if it is the round that counts.

In each round of this task, you will buy up to two different fictitious "fruits" at up to two "stores". You have a separate gift certificate (denominated in experimental currency units – ECUs) at each store that can be spent only at that store. However, your monetary earnings for the experiment are based on the total amount of each fruit in your final bundle for a round after you have completed your shopping at all stores.

The first page of each round will announce the number of stores in that round. At the end of each round, raise your hand so that the experimenter can collect your decisions and give you the decision sheet for the next round.

How Your Payment is Determined

Your monetary payment will be calculated from your final bundle in the round that counts according to the function

$$Payment = \frac{2}{5}\left(\sqrt{\#apples} + \sqrt{\#oranges}\right)^2.$$

To help you calculate the payment you would receive for a final bundle, we have provided tables at the end of the experiment that indicates the payment that would result from all possible final bundles (and some impossible ones).

As an example of how your payment will be calculated, suppose you buy:

1 apple and 5 oranges at Store 1,

2 apples and 6 oranges at Store 2.

Then your final bundle is

3 apples and 11 oranges.

To calculate your payment locate the entry in the "3 apples" column and the "11 oranges" row of the payment table.

Notice three features of the payment table:

(i) A final bundle with more of every fruit earns a higher payment.
(ii) A mix of fruits earns a higher payment: a final bundle with 5 apples and 5 oranges earns you a higher payment than a final bundle with 8 apples and 2 oranges, which in turn earns a higher payment than a final bundle with 10 apples and 0 oranges.
(iii) A final bundle with 7 apples and 3 oranges earns the same final payment as a final bundle with 3 apples and 7 oranges.

If the prices of apples and oranges are not the same, you thus face a trade-off between buying as many units of fruit as possible versus buying a mix that includes both fruits.

Figure 16. Shopping: Quiz

How to Shop in each Store

You will have a separate gift certificate at each store denominated in Experimental Currency Units (ECUs). The page for each store will present you with the prices of the fruits in that store. You must highlight one of the feasible apple-orange-watermelon combinations at each store to spend your gift certificate. Feasible combinations will be denoted in a list. If that combination does not appear in the, then it is not affordable with your gift certificate at that store.

To illustrate how you make your decision in each store, consider the following hypothetical store; you have a 6 ECU gift certificate for this store, and apples and oranges each cost 1 ECU per unit of fruit. Then your store page will be laid out as follows.

Store

You have a **6 ECU** gift certificate to spend.

The price of apples is **1 ECU per apple**.

The price of oranges is **1 ECU per orange**.

Please highlight a feasible combination of apples and oranges from the list below to make your purchase from this store.

0 apples and 6 oranges.

1 apple and 5 oranges.

2 apples and 4 oranges.

3 apples and 3 oranges.

4 apples and 2 oranges.

5 apples and 1 orange.

6 apples and 0 oranges.

Question 1.

How much would you earn if a round with only the store above was the round that counts, and you had chosen the bundle you indicated above?

Question 2.

Suppose that the round that counts had two stores. In Store 1, you bought 1 apple and 4 oranges. In store 2, you bought 3 apples and 5 oranges. What would your earnings be for the experiment?

Figure 17. Shopping: D3.2 Decision Sheet

Store 2

You have a **24 ECU** gift certificate at Store 2.

The price of apples is **3 ECU per apple**.

The price of oranges is **2 ECU per orange**.

Please highlight a feasible combination of apples and oranges from the list below to make your purchase from this store.

0 apples and 12 oranges.

0 apples and 11 oranges.

1 apple and 10 oranges.

2 apples and 9 oranges.

2 apples and 8 oranges.

3 apples and 7 oranges.

4 apples and 6 oranges.

4 apples and 5 oranges.

5 apples and 4 oranges.

6 apples and 3 oranges.

6 apples and 2 oranges.

7 apples and 1 orange.

8 apples and 0 oranges.

FIGURE 18. Shopping Payoff Table

	0 apples	1 apple	2 apples	3 apples	4 apples	5 apples	6 apples	7 apples	8 apples	9 apples	10 apples	11 apples
0 oranges	0.00	0.40	0.80	1.20	1.60	2.00	2.40	2.80	3.20	3.60	4.00	4.40
1 orange	0.40	1.60	2.33	2.99	3.60	4.19	4.76	5.32	5.86	6.40	6.93	7.45
2 oranges	0.80	2.33	3.20	3.96	4.66	5.33	5.97	6.59	7.20	7.79	8.38	8.95
3 oranges	1.20	2.99	3.96	4.80	5.57	6.30	6.99	7.67	8.32	8.96	9.58	10.20
4 oranges	1.60	3.60	4.66	5.57	6.40	7.18	7.92	8.63	9.33	10.00	10.66	11.31
5 oranges	2.00	4.19	5.33	6.30	7.18	8.00	8.78	9.53	10.26	10.97	11.66	12.33
6 oranges	2.40	4.76	5.97	6.99	7.92	8.78	9.60	10.38	11.14	11.88	12.60	13.30
7 oranges	2.80	5.32	6.59	7.67	8.63	9.53	10.38	11.20	11.99	12.75	13.49	14.22
8 oranges	3.20	5.86	7.20	8.32	9.33	10.26	11.14	11.99	12.80	13.59	14.36	15.10
9 oranges	3.60	6.40	7.79	8.96	10.00	10.97	11.88	12.75	13.59	14.40	15.19	15.96
10 oranges	4.00	6.93	8.38	9.58	10.66	11.66	12.60	13.49	14.36	15.19	16.00	16.79
11 oranges	4.40	7.45	8.95	10.20	11.31	12.33	13.30	14.22	15.10	15.96	16.79	17.60
12 oranges	4.80	7.97	9.52	10.80	11.94	13.00	13.99	14.93	15.84	16.71	17.56	18.39
13 oranges	5.20	8.48	10.08	11.40	12.57	13.65	14.67	15.63	16.56	17.45	18.32	19.17
14 oranges	5.60	8.99	10.63	11.98	13.19	14.29	15.33	16.32	17.27	18.18	19.07	19.93
15 oranges	6.00	9.50	11.18	12.57	13.80	14.93	15.99	17.00	17.96	18.90	19.80	20.68
16 oranges	6.40	10.00	11.73	13.14	14.40	15.56	16.64	17.67	18.65	19.60	20.52	21.41
17 oranges	6.80	10.50	12.26	13.71	15.00	16.18	17.28	18.33	19.33	20.30	21.23	22.14
18 oranges	7.20	10.99	12.80	14.28	15.59	16.79	17.91	18.98	20.00	20.98	21.93	22.86
19 oranges	7.60	11.49	13.33	14.84	16.17	17.40	18.54	19.63	20.66	21.66	22.63	23.57
20 oranges	8.00	11.98	13.86	15.40	16.76	18.00	19.16	20.27	21.32	22.33	23.31	24.27
21 oranges	8.40	12.47	14.38	15.95	17.33	18.60	19.78	20.90	21.97	23.00	23.99	24.96
22 oranges	8.80	12.95	14.91	16.50	17.90	19.19	20.39	21.53	22.61	23.66	24.67	25.65

	12 apples	13 apples	14 apples	15 apples	16 apples	17 apples	18 apples	19 apples	20 apples	21 apples	22 apples
0 oranges	4.80	5.20	5.60	6.00	6.40	6.80	7.20	7.60	8.00	8.40	8.80
1 orange	7.97	8.48	8.99	9.50	10.00	10.50	10.99	11.49	11.98	12.47	12.95
2 oranges	9.52	10.08	10.63	11.18	11.73	12.26	12.80	13.33	13.86	14.38	14.91
3 oranges	10.80	11.40	11.98	12.57	13.14	13.71	14.28	14.84	15.40	15.95	16.50
4 oranges	11.94	12.57	13.19	13.80	14.40	15.00	15.59	16.17	16.76	17.33	17.90
5 oranges	13.00	13.65	14.29	14.93	15.56	16.18	16.79	17.40	18.00	18.60	19.19
6 oranges	13.99	14.67	15.33	15.99	16.64	17.28	17.91	18.54	19.16	19.78	20.39
7 oranges	14.93	15.63	16.32	17.00	17.67	18.33	18.98	19.63	20.27	20.90	21.53
8 oranges	15.84	16.56	17.27	17.96	18.65	19.33	20.00	20.66	21.32	21.97	22.61
9 oranges	16.71	17.45	18.18	18.90	19.60	20.30	20.98	21.66	22.33	23.00	23.66
10 oranges	17.56	18.32	19.07	19.80	20.52	21.23	21.93	22.63	23.31	23.99	24.67
11 oranges	18.39	19.17	19.93	20.68	21.41	22.14	22.86	23.57	24.27	24.96	25.65
12 oranges	19.20	19.99	20.77	21.53	22.29	23.03	23.76	24.48	25.19	25.90	26.60
13 oranges	19.99	20.80	21.59	22.37	23.14	23.89	24.64	25.37	26.10	26.82	27.53
14 oranges	20.77	21.59	22.40	23.19	23.97	24.74	25.50	26.25	26.99	27.72	28.44
15 oranges	21.53	22.37	23.19	24.00	24.79	25.57	26.35	27.11	27.86	28.60	29.33
16 oranges	22.29	23.14	23.97	24.79	25.60	26.39	27.18	27.95	28.71	29.46	30.21
17 oranges	23.03	23.89	24.74	25.57	26.39	27.20	27.99	28.78	29.55	30.32	31.07
18 oranges	23.76	24.64	25.50	26.35	27.18	27.99	28.80	29.59	30.38	31.15	31.92
19 oranges	24.48	25.37	26.25	27.11	27.95	28.78	29.59	30.40	31.19	31.98	32.76
20 oranges	25.19	26.10	26.99	27.86	28.71	29.55	30.38	31.19	32.00	32.80	33.58
21 oranges	25.90	26.82	27.72	28.60	29.46	30.32	31.15	31.98	32.80	33.60	34.40
22 oranges	26.60	27.53	28.44	29.33	30.21	31.07	31.92	32.76	33.58	34.40	35.20

Appendix I. Online Risk Experiment Results

# errors	Tabs, Basic 0	1	2	Side-by-Side, Basic 0	1	2	Tabs, Examine 0	1	2	Side-by-Side, Examine 0	1	2
NB-WARP (D1.1 and D5)	41	50	52	36	41	43	26	34	38	26	38	45
NB-WARP (D1.2 and D4))	43	46	47	38	43	43	34	37	39	38	42	46
NB-WARP (D3.2 and D5)	41	49	53	26	34	43	25	34	41	25	40	45
NB-WARP (D1.1 and D3.2)	40	49	53	30	37	46	27	38	44	23	39	41
NB-WARP (all)	29	36	41	21	25	36	18	28	33	16	29	36
BB-WARP (D1 and D2)	5	6	49	8	8	48	4	5	39	3	5	43
BB-Mon (D1)	2	3	6	2	2	3	5	5	8	4	4	6
BB-Mon (D3)	4	5	6	1	1	3	4	4	5	0	0	4
BB-Mon (both)	2	2	3	1	1	1	4	4	4	0	0	0
# subjects		56			50			46			48	

Entries count the # of subjects who pass test at the listed error allowance.

Table 17. Tests of NB-WARP and BB-WARP

# errors	Tabs, Basic 0	1	2	Side-by-Side, Basic 0	1	2	Tabs, Examine 0	1	2	Side-by-Side, Examine 0	1	2
NB-SARP	24	29	33	18	21	27	17	24	27	13	22	26
BB-SARP	0	0	0	0	0	0	1	1	2	0	0	0
PNB	31	35	41	26	31	34	27	30	33	28	32	37
PNB-PE	32	35	41	26	32	35	28	31	35	29	32	38
# subjects		56			50			46			48	

Entries count the # of subjects who pass each test at the listed error allowance.

Table 18. Full Tests of Symmetric Models

	Percent Selten Score Maximized			
	Tabs, Basic	Side-by-Side, Basic	Tabs, Examine	Side-by-Side, Examine
Broad Bracketing	0	0	2	0
Narrow Bracketing	48	41	34	45
PNB	1	0	0	0
PNB-PE	0	1	1	0
Unclassified	7	8	9	3
# subjects	56	50	46	48

Table 19. Classification of subjects

Appendix J. Online Shopping Experiment Results

	D1			D3			Both			Full		
# errors	0	1	2	0	1	2	0	1	2	0	1	2
NB	6	7	20	18	23	30	4	5	11	4	5	8
BB	5	5	7	4	4	4	3	3	3	3	3	3
PNB	12	15	33	23	29	37	7	9	15	7	9	12
PNB-PE	12	15	33	23	29	37	7	9	15	7	9	12
# subjects						46						

Entries count the # of subjects who pass each test at the listed error allowance.

Table 20. Shopping Tests

	Percent Selten Score Maximized Online Shopping
Broad Bracketing	5
Narrow Bracketing	33
PNB	4
PNB-PE	0
Unclassified	4
# of subjects	46

Table 21. Classification of subjects

Versions 2 and 3. When we conducted our main Online Shopping Experiment, we recruited 150 US-based subjects from Prolific Academic, randomly assigning each to one of the three treatments, with 46 assigned to the main treatment, 54 to V2, and 49 in V3. Treatment V1 is identical to the Paper one reported, without a kink in the budget of D2. The two other versions of our Online Shopping Experiment were intended to reduce the reliance of our tests on corner solutions being optimal for a broad bracketer. In V2, we made D1.1 and D3.1 binary choices that forces each subject to pick a corner or near-corner solution (Table 22). In V3, we used the full budget set consistent with the implied prices in V2 for D1.1 and D3.1. In order to be able to distinguish between the predictions of narrow and broad bracketing, we multiplied all budgets by a factor of 4 and divided the payoff function by 4 (to retain the same stakes). Table 22 presents the budget sets used in each version.

		V1			V2			V3		
Decision	Part	I	p_a	p_o	I	p_a	p_o	I	p_a	p_o
D1	1	8	2	1	36	1 (sold only in a 30-pack)	6	36	1	6
	2	24	2	2	108	3	4	108	3	4
D2	1	40	2	2	160	2	2	160	2	2
D3	1	30	3	3	24	1 (sold only in a 24-pack)	4	24	1	4
	2	24	3	2	48	2	3	48	2	3
D4	1	12	1	1	48	1	1	48	1	1
D5	1	48	6	4	192	6	4	192	6	4

p_a: price/apple
p_o: price/orange

Table 22. Experimental Tasks for Online Shopping Experiments V1, V2, and V3

Unfortunately, these changes and the move to online made behavior much noisier, to the point that it is impossible to draw many conclusions from V2 and V3. In hindsight, we believe that enriching the budget space made it too difficult for a subject to efficiently use the calculator to explore the budget space, which in turn increased noise. This is seen most clearly in D5, where the optimal allocation in the original experiment and V1 is to purchase 4 apples and 6 oranges, and the optimum in V2 and V3 is to purchase 12 apples and 30 oranges (discreteness). In the original paper version, only 9% of subjects fail to make the optimal purchase, and on average, subjects are 0.20 oranges off the optimal allocation. In V1, 42%

of subjects make a mistake, being off by 1.16 oranges on average. In V2 and V3, more than 87% made a mistake and more than half were off by more than 9 oranges. Table 23 reports the fraction of optimal choices and the median and mean errors from payoff-maximization in D4 and D5.[46] These decisions are directly comparable across all versions of the experiment (noting that the budget is multiplied by 4 in V2 and V3). Strikingly, a far higher fraction of subjects make errors in all online versions, and this is especially pronounced in V2 and V3. The average magnitude of errors is also much larger in the online experiments than in the original paper experiment. This difference is consistent with previous work by Snowberg & Yariv (2021) that showed that student subjects tend to exhibit less noisy behavior than subjects recruited from a representative sample or from Mechanical Turk (which are more similar to the Prolific Academic subject pool).

	D4			D5		
	Fraction	Median	Mean	Fraction	Median	Mean
Paper	0.01	0	0.02	0.09	0	0.20
V1	0.13	0	0.27	0.41	0	1.16
V2	0.18	0	0.84	0.87	9	8.24
V3	0.14	0	0.53	0.90	9	8.16

Table reports the fraction of participants making an error, the median and mean number of oranges away from the optimum for the main Paper experiments and versions 1, 2, and 3 of the online experiments separately for D4 and D5.

TABLE 23. Comparing Errors

Our goal was to compare broad bracketing in V2 with broad bracketing in V3. While we attempted to perform our classification exercise with V2 and V3, the classification was much noisier. In V2, 31 of 54 subjects are unclassified, with 9 classified to broad, 5 to narrow, and 9 to partial. However, only 15 subjects were within 8 errors of any of the three models (4 classified broad, 3 narrow, and 8 partial). In V3, only 4 of 49 subjects could be classified (1 broad, 1 narrow, 2 partial). This appears to be primarily due to the higher level of noise in the online experiment. Unsurprisingly given the small sample, a Fisher's exact test of the

[46] Recall that we measure errors in terms of lines on our decision sheets in our original experiments and analogously in terms of available slider positions in our online experiments. In D4 and D5, there was one slider position for every affordable number of oranges.

hypothesis that there is no difference in the ratio of number of subjects classified as narrow to those classified as broad between V2 and V3 is insignificant, $p = 1.0$.

Table 24 reports the number of people in each treatment who are within 0, 4 and 8 errors of the optimal allocation for NB and BB. To deal with the high level of noise, we compare the ratio of people who pass NB versus BB across the two treatments. For each error allowance, we perform a separate Fisher's exact test of the hypothesis that there is no difference in the ratio of the number of subject within that allowance of narrow's prediction to the the number of subject within that allowance of broad's prediction between V2 and V3. Unsurprisingly given the small sample, we find no significant differences in the ratio between these two treatments. This further suggests that extreme avoidance is not driving our results.

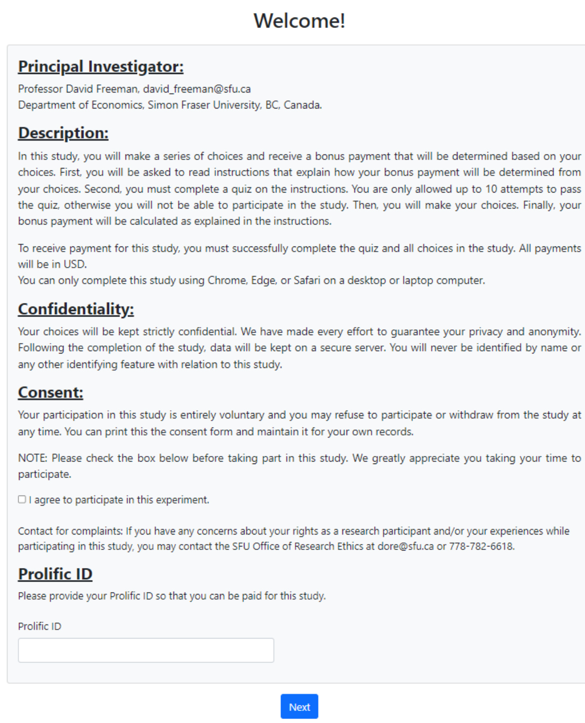
	V2		V3		V2 vs. V3
# errors	NB	BB	NB	BB	p-value
0	2	0	0	0	1.00
4	3	1	0	0	1.00
8	4	4	1	0	1.00

Each entry is the number of subjects who are within that error allowance of NB/BB.

Table 24. V2 and V3: testing the broad to narrow bracketer ratio

Appendix K. Online Risk Experiment Screenshots

The Online Risk Experiment and Online Shopping Experiment were both programmed using oTree (Chen *et al.*, 2016).



Welcome!

Principal Investigator:

Professor David Freeman, david_freeman@sfu.ca
Department of Economics, Simon Fraser University, BC, Canada.

Description:

In this study, you will make a series of choices and receive a bonus payment that will be determined based on your choices. First, you will be asked to read instructions that explain how your bonus payment will be determined from your choices. Second, you must complete a quiz on the instructions. You are only allowed up to 10 attempts to pass the quiz, otherwise you will not be able to participate in the study. Then, you will make your choices. Finally, your bonus payment will be calculated as explained in the instructions.

To receive payment for this study, you must successfully complete the quiz and all choices in the study. All payments will be in USD.
You can only complete this study using Chrome, Edge, or Safari on a desktop or laptop computer.

Confidentiality:

Your choices will be kept strictly confidential. We have made every effort to guarantee your privacy and anonymity. Following the completion of the study, data will be kept on a secure server. You will never be identified by name or any other identifying feature with relation to this study.

Consent:

Your participation in this study is entirely voluntary and you may refuse to participate or withdraw from the study at any time. You can print this the consent form and maintain it for your own records.

NOTE: Please check the box below before taking part in this study. We greatly appreciate you taking your time to participate.

☐ I agree to participate in this experiment.

Contact for complaints: If you have any concerns about your rights as a research participant and/or your experiences while participating in this study, you may contact the SFU Office of Research Ethics at dore@sfu.ca or 778-782-6618.

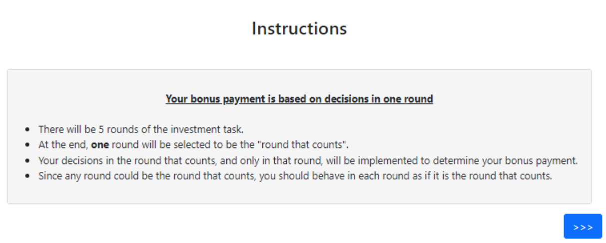
Prolific ID

Please provide your Prolific ID so that you can be paid for this study.

Prolific ID

Next

Figure 19. Consent Form



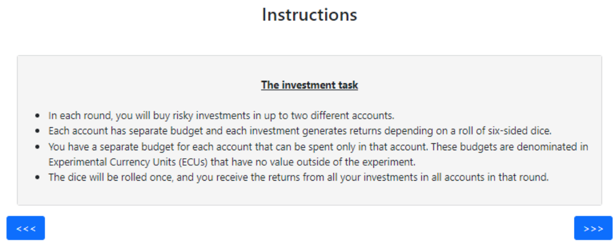
Instructions

Your bonus payment is based on decisions in one round

- There will be 5 rounds of the investment task.
- At the end, **one** round will be selected to be the "round that counts".
- Your decisions in the round that counts, and only in that round, will be implemented to determine your bonus payment.
- Since any round could be the round that counts, you should behave in each round as if it is the round that counts.

>>>

Figure 20. Instructions, Page 1



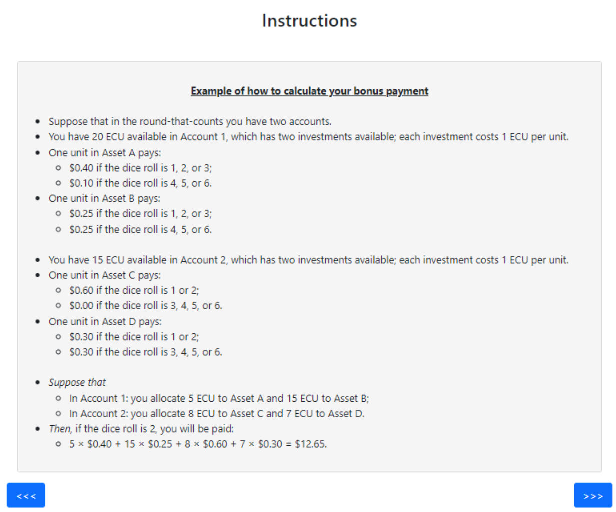
Instructions

The investment task

- In each round, you will buy risky investments in up to two different accounts.
- Each account has separate budget and each investment generates returns depending on a roll of six-sided dice.
- You have a separate budget for each account that can be spent only in that account. These budgets are denominated in Experimental Currency Units (ECUs) that have no value outside of the experiment.
- The dice will be rolled once, and you receive the returns from all your investments in all accounts in that round.

<<< >>>

FIGURE 21. Instructions, Page 2



Instructions

Example of how to calculate your bonus payment

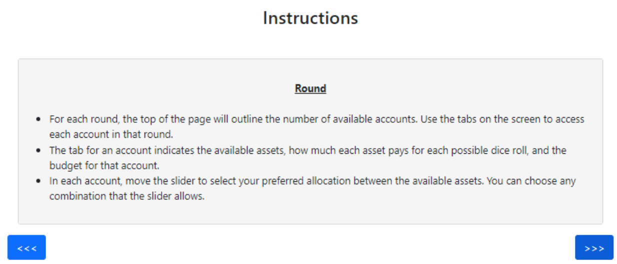
- Suppose that in the round-that-counts you have two accounts.
- You have 20 ECU available in Account 1, which has two investments available; each investment costs 1 ECU per unit.
- One unit in Asset A pays:
  - $0.40 if the dice roll is 1, 2, or 3;
  - $0.10 if the dice roll is 4, 5, or 6.
- One unit in Asset B pays:
  - $0.25 if the dice roll is 1, 2, or 3;
  - $0.25 if the dice roll is 4, 5, or 6.

- You have 15 ECU available in Account 2, which has two investments available; each investment costs 1 ECU per unit.
- One unit in Asset C pays:
  - $0.60 if the dice roll is 1 or 2;
  - $0.00 if the dice roll is 3, 4, 5, or 6.
- One unit in Asset D pays:
  - $0.30 if the dice roll is 1 or 2;
  - $0.30 if the dice roll is 3, 4, 5, or 6.

- *Suppose that*
  - In Account 1: you allocate 5 ECU to Asset A and 15 ECU to Asset B;
  - In Account 2: you allocate 8 ECU to Asset C and 7 ECU to Asset D.
- *Then*, if the dice roll is 2, you will be paid:
  - 5 × $0.40 + 15 × $0.25 + 8 × $0.60 + 7 × $0.30 = $12.65.

<<< >>>

FIGURE 22. Instructions, Page 3



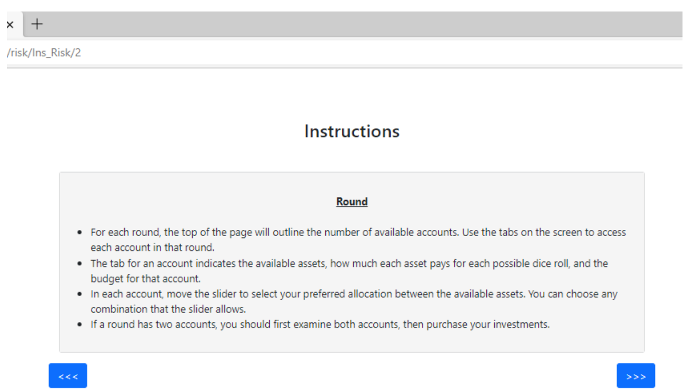
Instructions

Round

- For each round, the top of the page will outline the number of available accounts. Use the tabs on the screen to access each account in that round.
- The tab for an account indicates the available assets, how much each asset pays for each possible dice roll, and the budget for that account.
- In each account, move the slider to select your preferred allocation between the available assets. You can choose any combination that the slider allows.

<<< >>>

FIGURE 23. Instructions, Page 4, Basic Versions



/risk/Ins_Risk/2

Instructions

Round

- For each round, the top of the page will outline the number of available accounts. Use the tabs on the screen to access each account in that round.
- The tab for an account indicates the available assets, how much each asset pays for each possible dice roll, and the budget for that account.
- In each account, move the slider to select your preferred allocation between the available assets. You can choose any combination that the slider allows.
- If a round has two accounts, you should first examine both accounts, then purchase your investments.

<<< >>>

FIGURE 24. Instructions, Page 4, Examine Versions

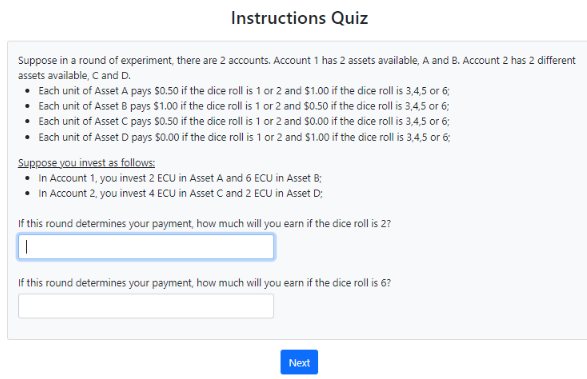


FIGURE 25. Quiz, Page 1, Basic Versions

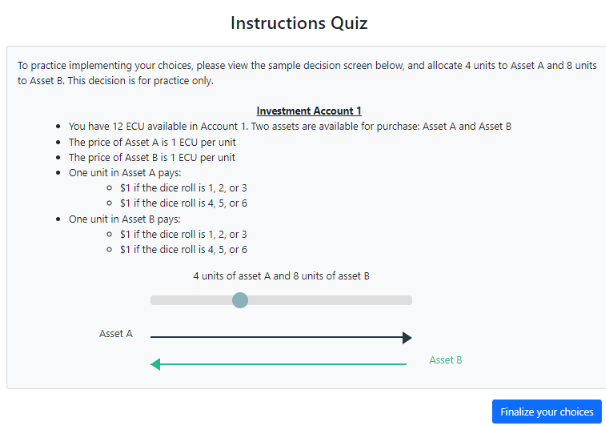


FIGURE 26. Quiz, Page 2, Basic Versions

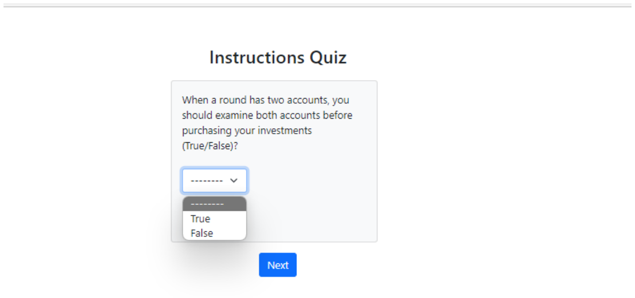


FIGURE 27. Quiz, Additional Question, Examine Versions

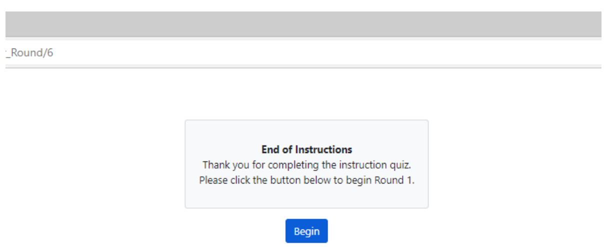


FIGURE 28. Quiz End

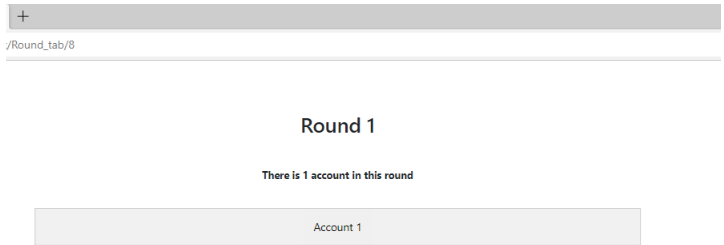


FIGURE 29. Round 1 initial screen, one-part decision randomly selected for Round 1

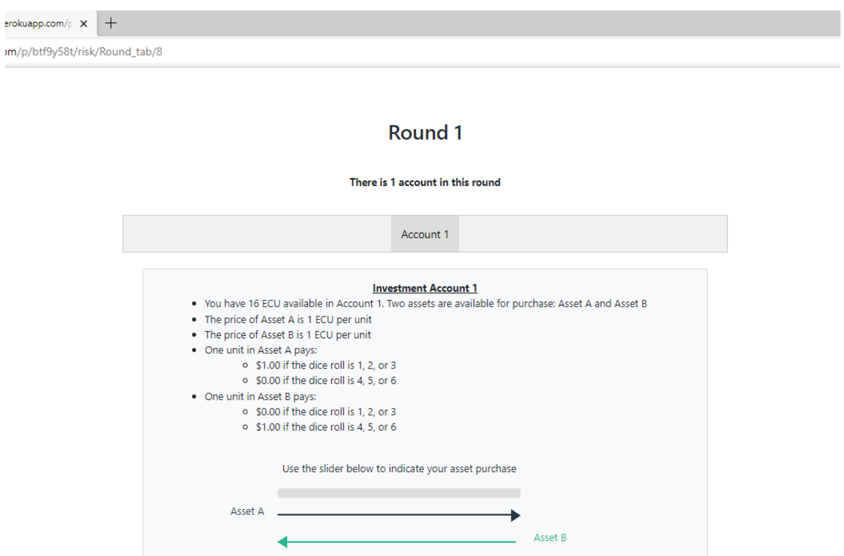


FIGURE 30. Round 1 decision screen, one-part decision randomly selected for Round 1

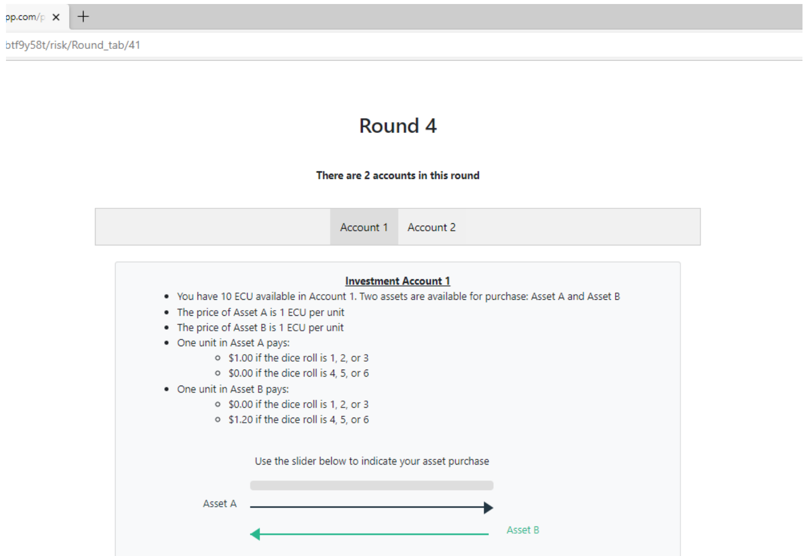


FIGURE 31. Round 4 decision screen, tabs version, two-part decision randomly selected for Round 4

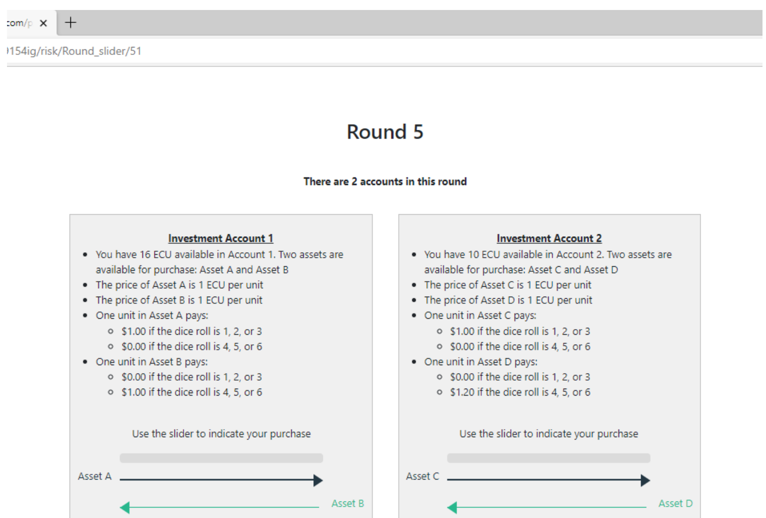


FIGURE 32. Round 5 decision screen, side-by-side version, two-part decision randomly selected for Round 5

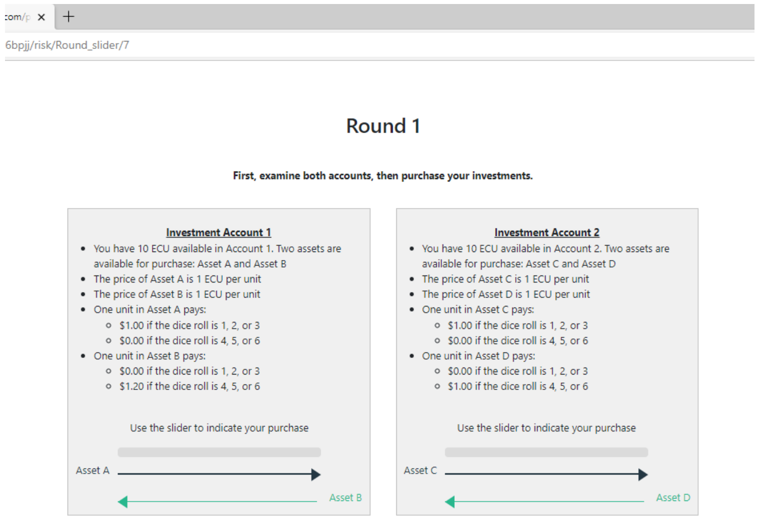


FIGURE 33. Round 1 decision screen, examine + side-by-side version, two-part decision randomly selected for Round 1

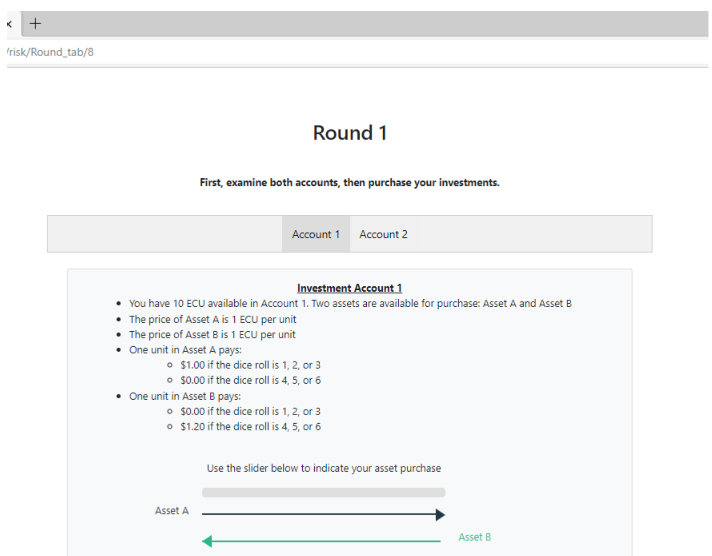


FIGURE 34. Round 1 decision screen, examine + tabs version, two-part decision randomly selected for Round 1

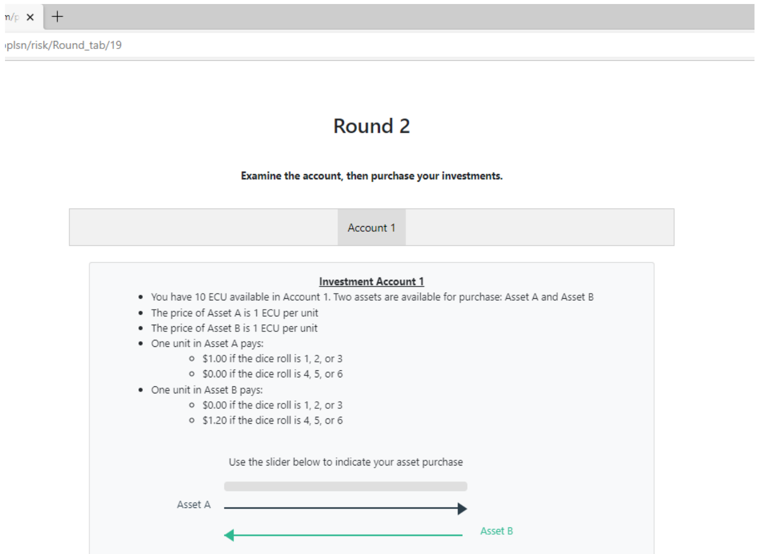


FIGURE 35. Round 2 decision screen, examine + tabs version, one-part decision randomly selected for Round 2

The final survey was the same as in the Online Shopping Experiment, and screenshots of it are provided in the next section.

Appendix L. Online Shopping Experiment Screenshots

The recruitment procedure and consent form were identical to the Online Risk Experiment.

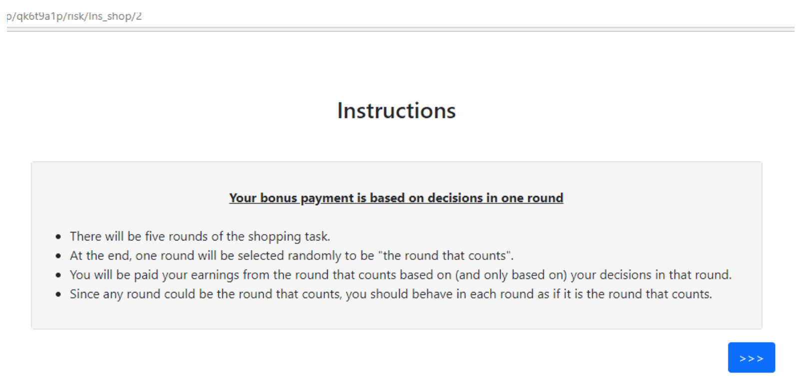


Figure 36. Instructions Page 1

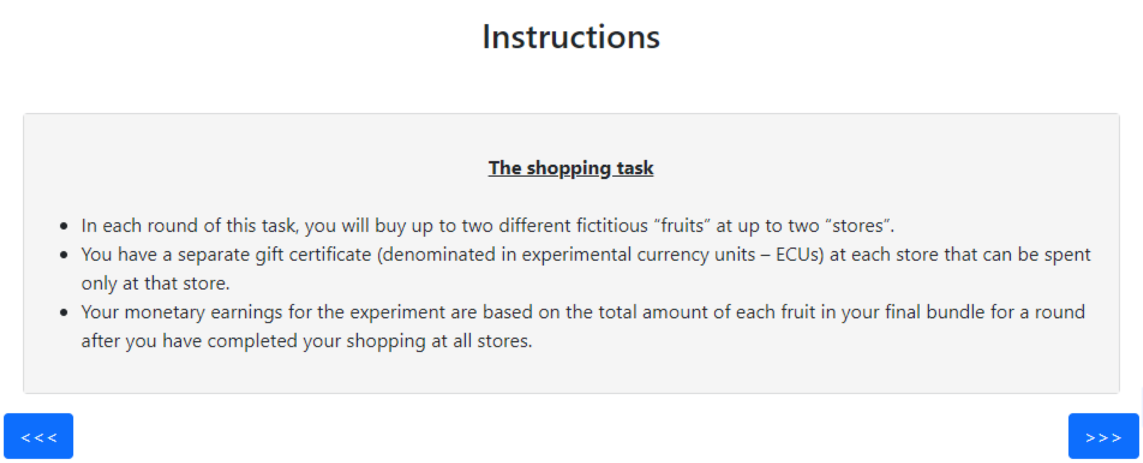


Figure 37. Instructions Page 2

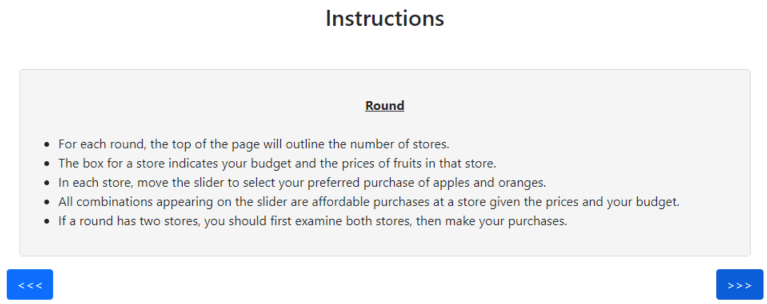


FIGURE 38. Instructions Page 3

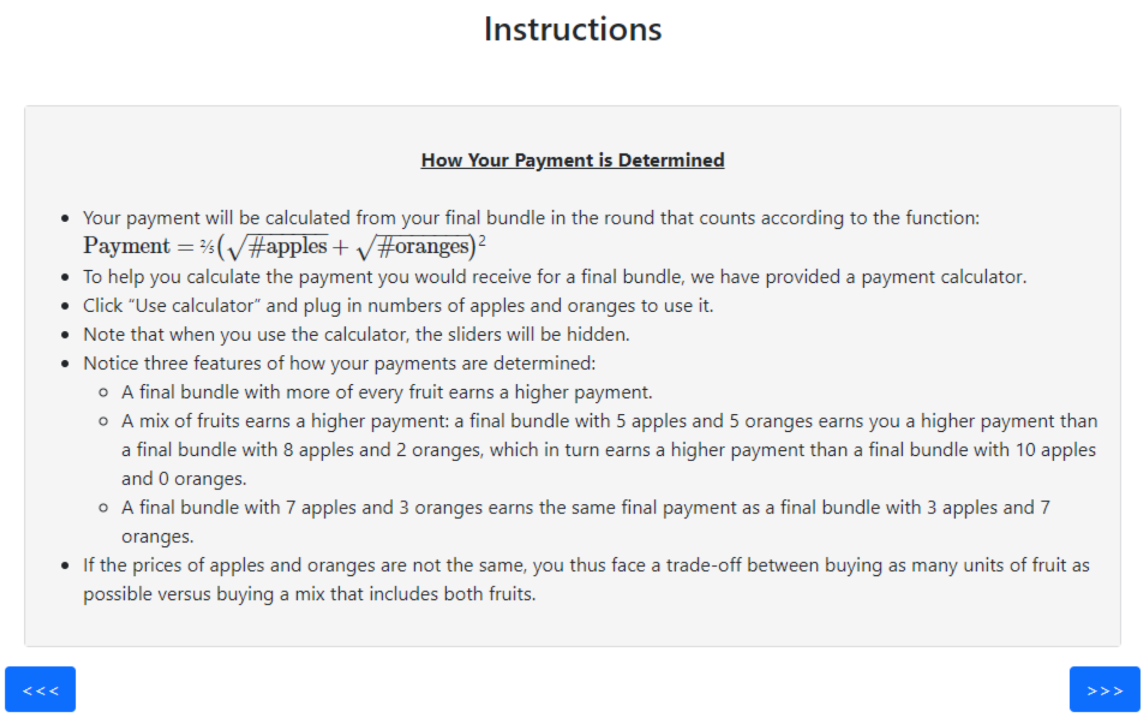


FIGURE 39. Instructions Page 4

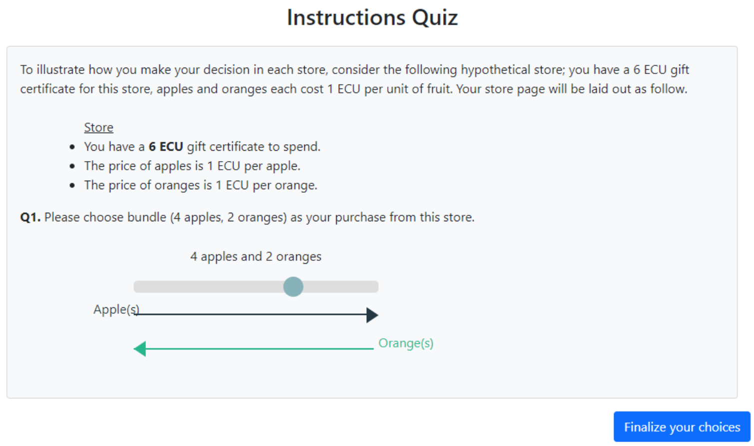


FIGURE 40. Instructions Quiz Q1

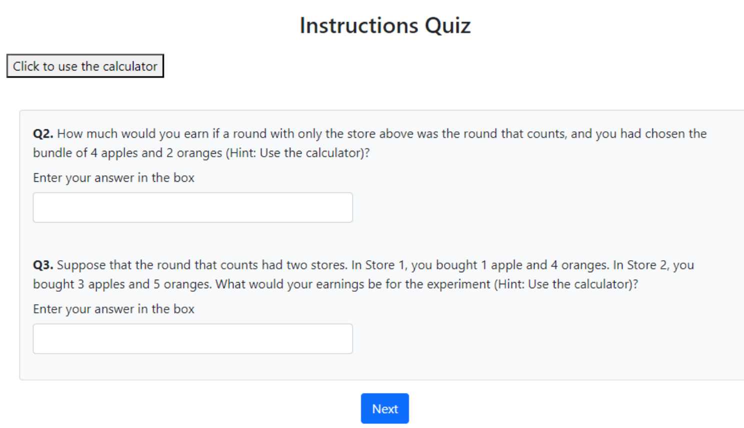


FIGURE 41. Instructions Quiz Q2

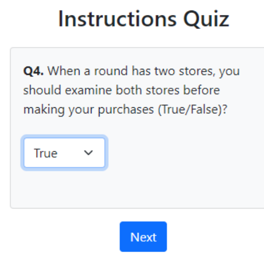


FIGURE 42. Instructions Quiz Q3

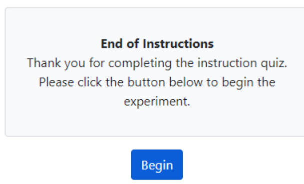


FIGURE 43. End of Instructions

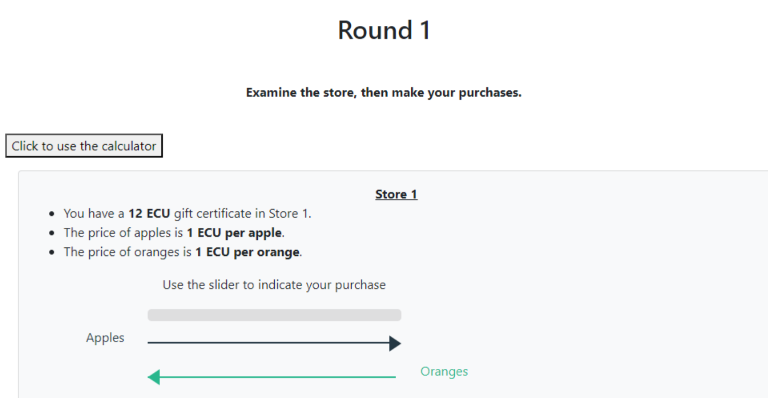


FIGURE 44. Sample Round 1 (randomly determined to be D4 in this case)

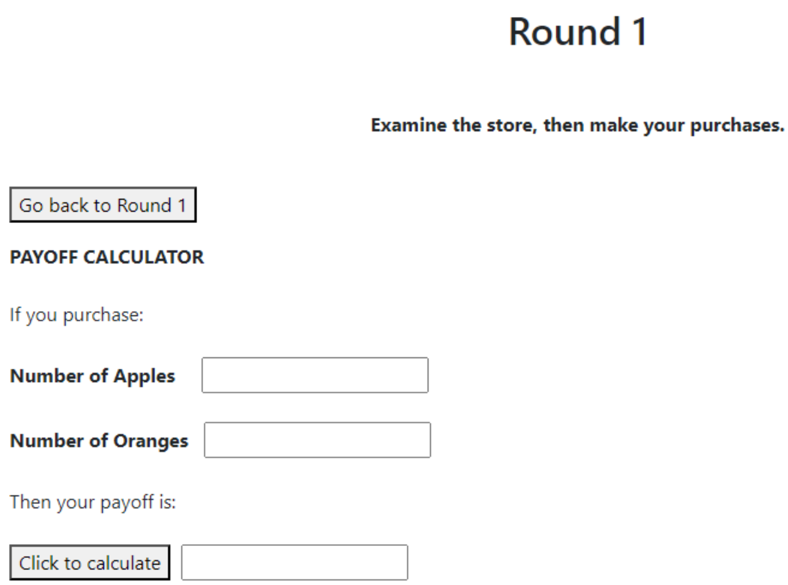


FIGURE 45. Calculator Screen

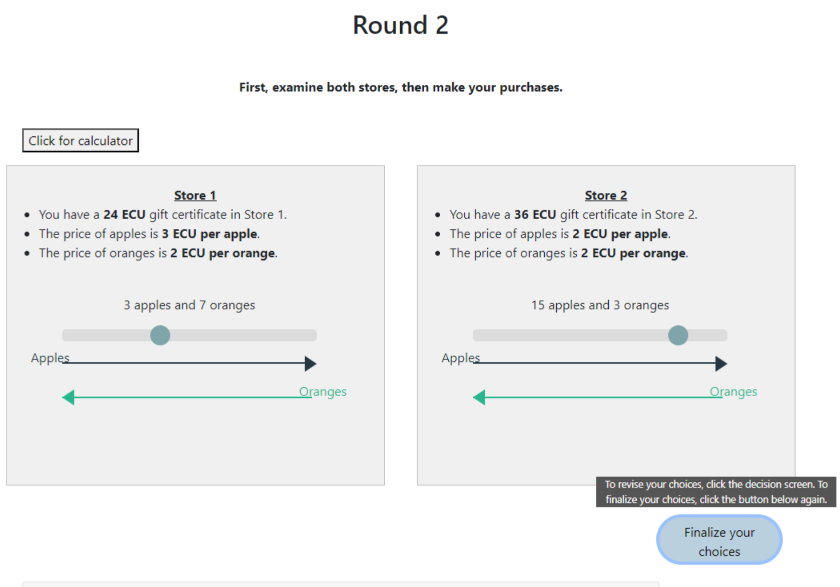


FIGURE 46. Sample Round 2 (randomly determined to be D3 in this case), showing prompt before finalizing choices

In order to simplify the visual interface and the programming task, our Online Shopping Experiment no longer offers a "sale" price in D2. Instead, it offers a budget of 40 ECU and prices of 2 ECU per fruit for both apples and oranges. This is visually shown in Figure 47.

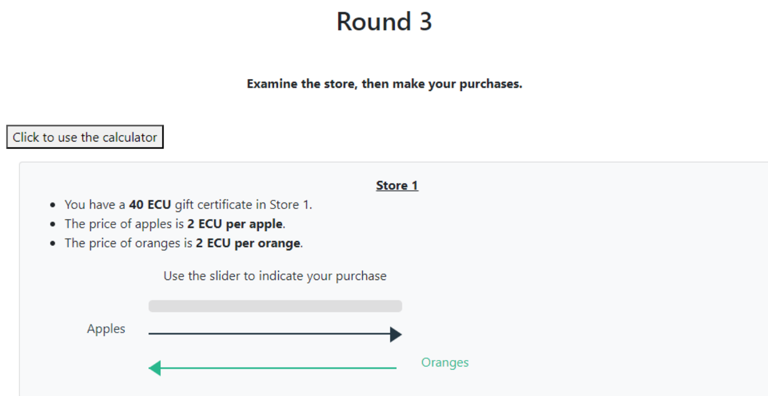


FIGURE 47. Sample Round 3 (randomly determined to be D2 in this case).

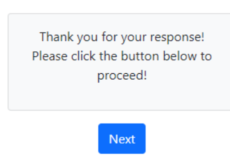


FIGURE 48. After Round Screen, appears after each decision screen in all online experiments

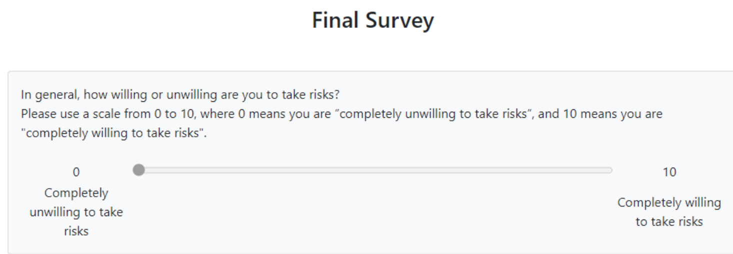


FIGURE 49. Survey Page 1 (self-reported risk aversion)

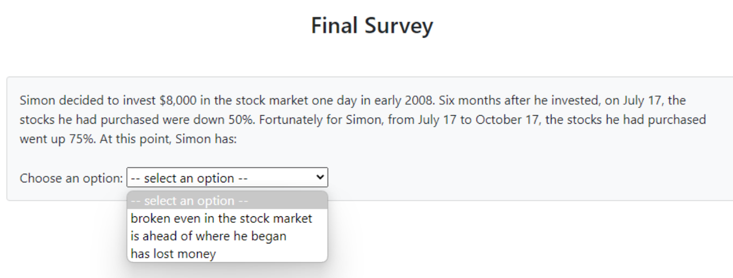


FIGURE 50. Survey Page 2 (CRT)

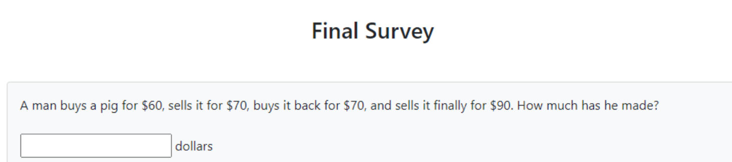
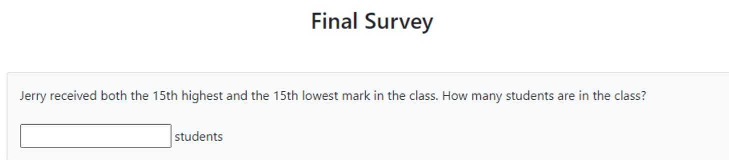


FIGURE 51. Survey Page 3 (CRT)

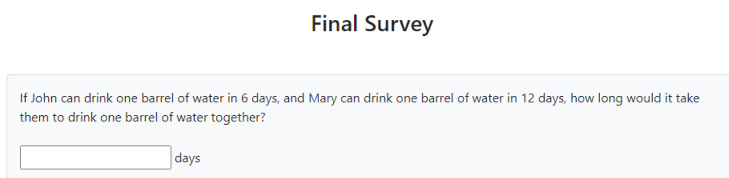


Final Survey

Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class?

students

FIGURE 52. Survey Page 4 (CRT)

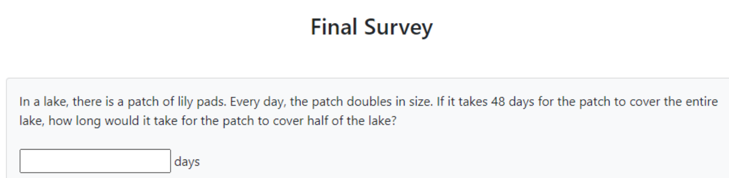


Final Survey

If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together?

days

FIGURE 53. Survey Page 5 (CRT)

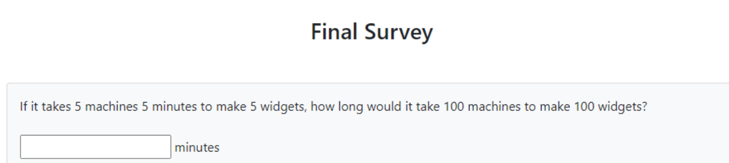


Final Survey

In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

days

FIGURE 54. Survey Page 6 (CRT)

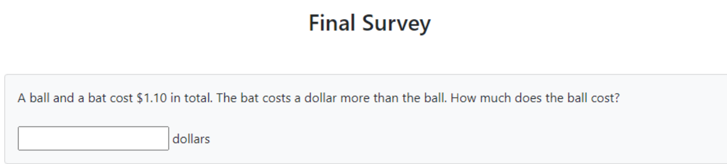


Final Survey

If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?

minutes

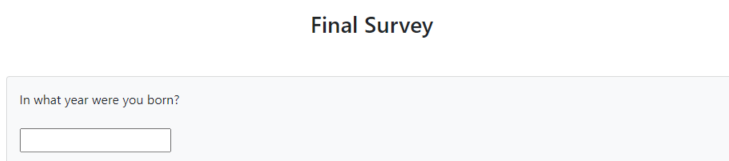
FIGURE 55. Survey Page 7 (CRT)


Final Survey

A ball and a bat cost $1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost?

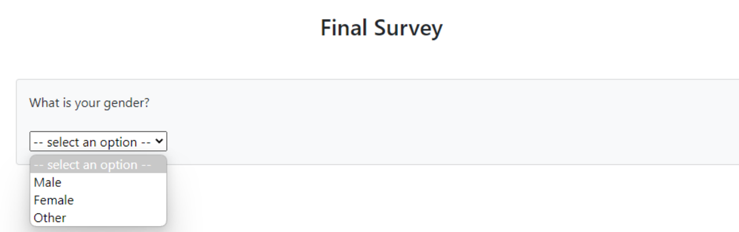
dollars

FIGURE 56. Survey Page 8 (CRT)


Final Survey

In what year were you born?

FIGURE 57. Survey Page 9 (age)

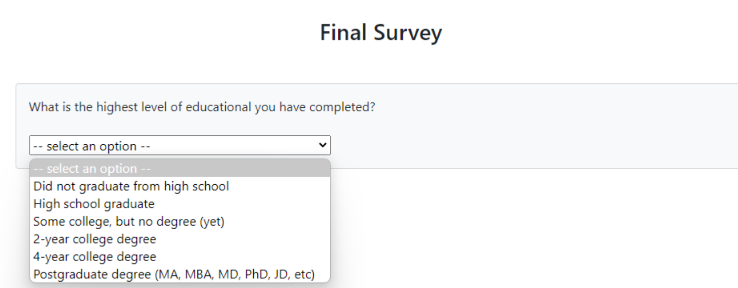

Final Survey

What is your gender?

-- select an option --

-- select an option --
Male
Female
Other

FIGURE 58. Survey Page 10 (gender)


Final Survey

What is the highest level of educational you have completed?

-- select an option --

-- select an option --
Did not graduate from high school
High school graduate
Some college, but no degree (yet)
2-year college degree
4-year college degree
Postgraduate degree (MA, MBA, MD, PhD, JD, etc)

FIGURE 59. Survey Page 11 (education)

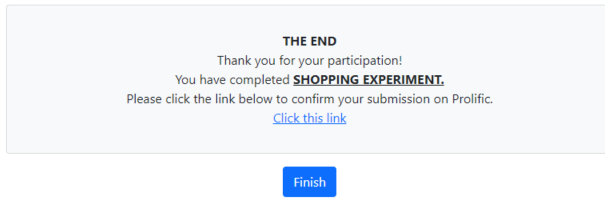


FIGURE 60. End Page