Rankings-Dependent Preferences: A Real Goods Matching Experiment

Andrew Kloosterman Peter Troyan

October 2022

Abstract

We investigate whether preferences for objects received via a matching mechanism are influenced by how highly agents rank them in their reported rank order list. We hypothesize that all else equal, agents receive greater utility for the same object when they rank it higher. The addition of rankings-dependent utility implies that it may not be a dominant strategy to submit truthful preferences to a strategyproof mechanism, and that non-strategyproof mechanisms that give more agents objects they *report* as higher ranked may increase market welfare. We test these hypotheses with a matching experiment in a strategyproof mechanism, the random serial dictatorship, and a non-strategyproof mechanism, the Boston mechanism. A novel feature of our experimental design is that the objects allocated in the matching markets are real goods, which allows us to directly measure rankings-dependence by eliciting values for goods both inside and outside of the mechanism. Our experimental results confirm that the elicited differences in values do decrease for lower-ranked goods. We find no differences between the two mechanisms for the rates of truth-telling and the final welfare.

1 Introduction

In strategyproof mechanisms, it is always an optimal strategy for agents to truthfully report their private information to the mechanism. This theoretical property is clearly appealing, as it gives a mechanism designer the ability to predict play and make meaningful statements about other criteria such as welfare. However, a growing body of empirical evidence has documented significant deviations from truthful reporting in such mechanisms. This issue is particularly important in matching markets, the focus in this paper, in which participants submit preference rankings of alternatives such as schools or medical residency programs to a centralized clearinghouse which determines the assignment. Evidence of non-truthful behavior in strategyproof matching mechanisms can be found both in the lab (Chen and Sönmez, 2006; Pais and Pintér, 2008; Li, 2017) and in high-stakes decisions in the field (Chen and Pereyra, 2019; Shorrer and Sóvágó, 2018; Hassidim et al., 2021).

Deviations from truthful reporting are harmful under the implicit assumption that agents' preferences are standard economic preferences in that the values for the objects they receive in the mechanism are determined solely by the characteristics of these objects. If this assumption holds, then, when we observe agents rank objects they value less above objects they value more in a strategyproof mechanism, we can claim that these deviations from truthful reporting are indeed "mistakes". But what if this underlying assumption is wrong? Then, these mistakes may not really be mistakes, but rather optimal behavior from agents with non-standard preferences.

In this paper, we explore the possibility that the rankings agents submit to the mechanism influence values. For instance, an agent may value an object higher when they rank it 2^{nd} compared to a counterfactual in which they receive the same object but rank it 4^{th}, because they suffer disutility when they get a low-ranked object. There are a number of reasons why receiving low-ranked objects may be undesirable. These include reference-dependent loss aversion where agents expect to get a high-ranked object and are disappointed when they do not (Dreyfuss et al., 2019; Meisner and von Wangenheim, 2021), ego utility where agents think that it looks good to others to receive a high-ranked object (Köszegi, 2006), preferences that focus on beating others rather than maximizing one's own utility (a "joy of winning", Cooper and Fang, 2008), or limited information on quality that instigates a 'curse of acceptance' whereby receiving a low-ranked object indicates that it is bad (Kloosterman and Troyan, 2020).

Following this discussion, we consider agents who have utility from receiving object x that takes the form

$$u(x) = v(x) + \rho(\text{rank}(x))$$

We call $v(x)$ the agent's *fundamental value* for object x; this corresponds to the standard economic preferences assumed in typical matching models. The second term, $\rho(\text{rank}(x))$, is an additional *rankings-dependent utility* component that is determined by how highly the agent ranked x in their reported preferences. Our main assumption is that $\rho(\cdot)$ is a decreasing function, i.e., all else equal, agents receive more utility when they rank an object higher.

Rankings-dependent utility has important consequences for real-world market design. First, with rankings dependence, deviations from truthful reporting may be optimal in a strategyproof mechanism because an agent may be able to attain a higher-ranked, though less desirable (in terms of fundamental value), object by ranking a less popular object highly.[1] Second, it has potential implications when evaluating the welfare of various mechanisms. Starting with the seminal paper of Abdulkadiroğlu and Sönmez (2003), there has been much written about the debate between strategyproof mechanism such as deferred acceptance (DA) and manipulable mechanisms such as the Boston mechanism (also called immediate acceptance). While DA gives better incentives to the agents, comparing the final rank distributions (how many students receive their reported first choice, reported second choice, etc.), which is a common outcome metric for many school districts, (unsurprisingly) shows better performance for mechanisms such as Boston which are designed with this goal in mind.[2]

The standard critique is that these data cannot be taken at face value, because the Boston mechanism gives clear incentives for agents to manipulate their preferences, even in the absence of rankings dependence (see Dur et al. (2018) for evidence of such behavior in a real-world school choice environment). However, if preferences are indeed rankings-dependent, this gives a

[1]The term truthful is not as apt with rankings-dependent utility. However, it is such standard usage in the matching literature without rankings dependence, in particular when discussing truthful or 'strategyproof' mechanisms that we investigate, that we continue to use this terminology to describe reporting preferences that rank the goods in the order of their fundamental value.

[2]Featherstone (2020) provides a detailed analysis of a general class of *rank-efficient mechanisms* that implement an assignment whose rank distribution is not first-order stochastically dominated by any other. See also Ortega and Klein (2022) and Troyan (2022), who look at the welfare and incentive properties of these mechanisms.

stronger argument for mechanisms in which more agents receive objects they report as higher-ranked: if more agents are receiving their reported first choices, and this in turn gives agents additional rankings-dependent utility, then such mechanisms can lead to increased total welfare. Indeed, in discussions with school district administrators, this is one argument that has been given for continued use of the Boston mechanism over strategyproof alternatives: parents just do not like to get something they ranked low in their list (Cambridge, MA School District, personal communication).

Thus, rankings-dependence has potentially important consequences for matching market design, and so determining whether these factors are relevant in practice is crucial. This is the main contribution of our paper: we design and implement a laboratory experiment to test the hypothesis of rankings-dependent utility. In our experiment, participants play in a matching market with five agents and five goods. As in most real-world implementations of matching markets, they play a one-shot game in which they are asked to submit a rank-order list of the five goods, and a mechanism is used to determine the final allocation to each agent.

The mechanisms that we use for our experiment are the random serial dictatorship (RSD) and the Boston mechanism. We chose these mechanisms because they are two canonical mechanisms that are widely used in practice. Further, RSD is strategyproof, while the Boston mechanism is not, yet the Boston mechanism may result in agents receiving higher-ranked goods. This allows us to answer not only our main question of rankings-dependent utility, but also to test the hypothesis that a non-strategyproof mechanism may be welfare-enhancing once rankings-dependent utility is taken into account.

We formalize hypotheses in Section 3 in a simplified environment that is relevant for the experiment. Participants have an incentive to rank less popular goods higher under both mechanisms to achieve a higher rankings-dependent utility; however, this incentive is stronger in Boston because popular goods are likely to be unavailable to them in later rounds of the mechanism. This means that if truthful reporting with respect to fundamental value is optimal in Boston, then it is also optimal in the RSD (Theorem 1), and that more agents get their top-ranked object in Boston than in RSD (Theorem 2). Finally, the extra incentive in Boston results in higher welfare in Boston than in the RSD in equilibrium (Theorem 3).

A novel and key feature of our experimental design is that we use real objects that are in the room at the time of the experiment and that the participants may take home with them. To determine whether utility is rankings-dependent, in Phase I of the experiment, we first elicit valuations for 20 common objects (backpacks, alarm clocks, phone chargers, etc.) with the multiple price list elicitation method. In Phase II of the experiment, five of the objects (a Fjallraven backpack, a Hydroflask water bottle, a Moleskine notebook, a generic ceramic coffee mug, and a package of 4 ballpoint pens) were chosen and the participants were asked to submit a rank-order list of these five objects to a mechanism. The mechanism (either RSD or Boston) produces an allocation of one object to each participant. After the mechanism, we once again elicit each participant's valuation for the object that they were allocated in the mechanism. The important feature of this design is that the Net Value (Phase II value minus Phase I value), which we abbreviate NV, measures $\rho(\text{rank}(x))$.

The use of real goods is a significant departure from the experimental literature on matching mechanisms which usually uses fictitious “goods” with induced monetary values (see Hakimov and Kübler (2021) for a recent survey of this literature). Measuring rankings-dependent utility would likely fail with induced values, because a participant’s elicited valuation for an amount of money is likely to just be that amount of money, and so $\rho(\text{rank}(x))$ would be zero. Further, using real goods makes the experiment more similar to real-world environments where participants must form their own preferences rather than have them being induced. It also mitigates a possible “experimenter demand” effect that could be present under an induced values framework, in which participants are given their preferences and asked to report them back. To our knowledge, we are the first to use real goods in a matching experiment, which is another contribution of our paper.

We present our results in Section 4. Our main hypothesis of interest is that the NV—which measures $\rho(\text{rank}(x))$—should be non-zero, and in particular decreasing in the reported rank of the good received in the mechanism. For the RSD treatment, we find clear evidence to support the hypothesis. The NV is nearly monotonically decreasing, from an average of +\$2.87 for participants who receive their top-ranked good to an average of −\$0.69 for participants who receive their fifth-ranked good (out of five). For the Boston treatment, there is a much smaller increase in NV for the first ranked good (only about +\$0.60), and, looking at the raw data averages, there is no clear evidence that NV is decreasing in rank. Non-parametric tests provide statistical support for these impressions. However, when we move to regression analysis to further explore these results, we find support for this hypothesis in both treatments, with the rank being a statistically significant predictor of NV for both the RSD treatment and the Boston treatment separately, as well as for the pooled sample. What appears to explain the discrepancy for Boston is the inclusion of a regressor for risk aversion, which was measured using the Holt-Laury switching point, and a regressor for the Phase I value. For Boston, risk aversion enters positively in the regression, indicating that all else equal, more risk averse individuals have larger increases in NV.

Next, we evaluate truth-telling with respect to fundamental values and welfare. Interestingly, we find no differences in the in the rates of truth-telling between the two treatments according to a wide range of measures. This is interesting given the plethora of experimental work going back to Chen and Sönmez (2006) that generally finds much less truth-telling in Boston than in deferred acceptance (which is equivalent to RSD in our setting), as predicted by the theory. These experiments use an induced values design, and details such as how much information participants are given about others’ preferences and their priorities are quite different from our set-up, where we do not have priorities and the use of real goods may make it difficult for participants to formulate beliefs about others’ preferences for the objects. Our results regarding truth-telling are in line with Pais and Pintér (2008) who find much higher rates of truth-telling in Boston when participants have little to no such information (they work with induced values rather than real goods). Ultimately, we find no differences in welfare between the two mechanisms, either. Perhaps some of this null finding is due to the goods we chose and/or the objects having no priority in our design. On the other hand, we do find that while the rates of truthful reporting are the same between the two treatments, the reasons underlying truthful reporting may be different. Using regression analysis, we find that non-truthful reporting in

RSD may be mistakes akin to the standard argument while non-truthful reporting in Boston is related to risk aversion and gender.

To summarize, we find that utility does appear to be rankings-dependent, although there are no differential effects for truth-telling between a strategyproof and a non-strategyproof mechanism. Our regression analysis indicates that this may be a combination of more mistakes in RSD that cancel out more strategic manipulation in Boston. While we also find no differences in welfare in our particular environment, the theory of rankings-dependent utility suggests that it can have welfare consequences. Having established the existence of rankings-dependent utility in a simple setting, more investigations are needed to determine the role it may play in the overall welfare of matching markets in other environments.

Related Literature

One of the main motivations for this paper is the growing body of evidence of "mistakes" in strategyproof mechanisms (see the first paragraph of the introduction for references). There are several ways one can proceed from these observations. Assuming that the underlying preference model is correct, and these mistakes are actually mistakes, one response is that the designer should simply invest more in communicating how the mechanism works and teaching the participants that truthful reporting is in their interest. Rees-Jones and Skowronek (2018) conduct a "lab-in-the-field" experiment in which they recruit medical students who have just gone through the National Resident Matching Program (NRMP) to participate in a related lab experiment, and find a significant fraction of participants did not report truthfully, despite having just participated in the same mechanism in a high stakes environment in which the NRMP invests heavily in tutorials describing how the mechanism works. If this approach is unsuccessful, another possibility is to design mechanisms that are more easily recognizable as strategyproof by the participants on their own. For instance, Li (2017) introduces the notion of obvious strategyproofness as a desideratum for mechanism design, with the idea being that if the mechanism is designed to satisfy this criterion, the players will be able to recognize themselves that truthful reporting is a dominant strategy, and will be more likely to report truthfully. Li's paper has led to a rapidly expanding literature on obvious mechanism design as a way to limit mistakes by participants.[3]

We take a different approach in this paper, which is a reassessment of the assumption that agent preferences are determined solely by the characteristics of the goods they receive. We discuss a few other recent papers that have explored related ideas.

The closest paper to ours theoretically is Meisner (2021). He proposes an equivalent model of utility that consists of both a fundamental value plus a rankings-dependent component. He then focuses on strategyproof mechanisms, and proves that any non-truthful preference ranking can be rationalized as optimal for some beliefs over match probabilities (what he refers to as "attainability distributions", which are determined by the mechanism itself combined with beliefs about the strategies of the other agents).

[3]Theoretical explorations include Ashlagi and Gonczarowski (2018), Troyan (2019), Pycia and Troyan (2020), and Bade and Gonczarowski (2017). For lab experiments, see Zhang and Levin (2017) and Bo and Hakimov (2019).

Dreyfuss et al. (2019), Dreyfuss et al. (2022), and Meisner and von Wangenheim (2021) all focus on expectations-based reference-dependent preferences (EBRD preferences for short, also referred to as EBLA for expectations-based loss aversion; Kőszegi and Rabin (2006)) as a possible explanation for seemingly dominated choices in strategyproof mechanisms. Dreyfuss et al. (2019) re-evaluate the experimental data of Li (2017), who finds mistakes in the non-obviously strategyproof RSD mechanism, and find that EBRD preferences might explain this behavior. Dreyfuss et al. (2022) study a lab experiment using four different implementations of the DA mechanism. Their experiment is very different from ours. Most notably, they use induced values and reduce the game to an individual decision problem in which participants know the probability of their priority score being above the threshold for admission to each school and are asked to choose actions that then induce lotteries over the schools.[4] The focus of their experiment is on the rates of non-truthful behavior (what they call non-straightforward behavior). They show that the variations in non-truthful behavior they find can be better rationalized by EBRD preferences compared to classical preferences. However, Dreyfuss et al. (2022) also write that "the EBRD model, while explaining a lot of the observed data, appears to be an incomplete explanation", and note that other explanations likely play an important role too.

Our experimental design, on the other hand, was devised to provide a direct measurement of rankings dependent utility, though we are agnostic on the underlying source of it. Additionally, we have participants play a multiplayer game, an area in which there is less work incorporating non-standard preferences relative to decision problems. Finally, while Dreyfuss et al. (2022) focus on explaining deviations from truthful behavior, we also consider the possible welfare implications of rankings-dependent preferences. Put together, we think these papers are all very complementary, and combined make a strong case that rankings-dependent preferences are likely to be relevant in matching market environments. Having established that, though, there is still much more work to be done to understand both the behavioral foundations and welfare implications of these findings.[5]

2 Model

2.1 Preferences and Mechanisms

There is a set $I = \{i_1, i_2, \ldots, i_N\}$ of agents and a set $X = \{x_1, x_2, \ldots, x_N\}$ of objects. A **matching** is a function $\mu : I \to X$ where $\mu(i) \in X$ is the object that is assigned to agent i. There are an equal number of agents and objects, and so we assume that all matchings assign a unique object to every agent, i.e., if $i \neq j$, then $\mu(i) \neq \mu(j)$. Let $\mathcal{M}$ denote the set of matchings. Notice that we assume each object has capacity 1, an equal number of agents and objects, and the objects have no preferences or priorities over the agents. All these restrictions can easily

[4] Dreyfuss et al. (2022) frame their experiment using the language of school choice and ask students to rank hypothetical schools which have an induced monetary value. The schools are analogous to the objects in our set-up.

[5] Meisner (2021) (discussed above) highlights some differences between EBRD and other sources of rankings-dependent utility, and proposes some ways to potentially differentiate between these alternatives in particular stylized environments.

be generalized to capture important features of real-world settings such as school choice. We make these assumptions because they are all that is needed to study our main phenomena of interest, and they ensure that the model corresponds as directly as possible to the experiment that we run.

Let $\mathcal{P}$ be the set of all strict ordinal rankings over X. For any $P_i \in \mathcal{P}$, we write xP_iy to denote that agent i strictly prefers x to y. We use R_i for the corresponding weak relation, i.e., xR_iy if either xP_iy or $x = y$. A **mechanism** is a function $\psi : \mathcal{P}^N \to \mathcal{M}$. We write $\psi_i(P_N) \in X$ for the object allocated to i at preference profile $P = (P_1, \ldots, P_N)$. Every mechanism induces a game in which the action space for each agent is $\mathcal{P}$.

In most of the matching literature, it is assumed that an agent's utility is determined solely by the object received. We deviate from this assumption by allowing an agent's utility to depend on both (i) the object received and (ii) the position in which they ranked the object in their preferences. Formally, agent i's utility from submitting a reported preference ranking P_i and receiving object x is

$$u_i(x, P_i) = v_i(x) + \rho(j). \tag{1}$$

where $j = |\{x' \in X : x'R_ix\}|$ is the rank of object x in the reported preference list. We call $v_i(x)$ agent i's **fundamental value** for object x, and $\rho(j)$ agent i's **rankings-dependent utility** from receiving the object that she ranked in the j^{th} position. Note that in this formulation, the function $\rho(\cdot)$ is the same for all agents. This could be generalized to allow for heterogeneity, but given our experimental design, it is infeasible to measure a different $\rho(j)$ for each agent and each j, and so we omit this generalization in the model.

The main assumption of our model is that, all else equal, participants will prefer to get objects when they rank them higher in their submitted preferences. For example, receiving good x provides more utility when it is ranked 2^{nd} than when it is ranked 4^{th}. Hence, we assume the function ρ is such that $\rho(1) \geq \rho(2) \geq \cdots \geq \rho(n)$. We discussed in the introduction possible microfoundations for rankings-dependent utility. In the experiment, we do not attempt to discern between these explanations, but seek simply to determine whether preferences do in fact depend on rankings.

For our experiment, we consider two mechanisms which we describe here: random serial dictatorship and the Boston mechanism.

Random Serial Dictatorship (RSD)

The random serial dictatorship mechanism works as follows. Each agent submits a strict ordinal ranking over all of the objects. The mechanism then draws an ordering of the agents randomly from the uniform distribution over all possible agent orderings, and proceeds in rounds as follows:

- Round 1: The first agent in the order is assigned their top-ranked object according to their submitted preference ranking.
- Round 2: The second agent in the order is assigned their top-ranked object that was not assigned in the first round.

- Round $k = 3, \ldots, N$: The k^{th} agent in the order is assigned their top-ranked object that was not assigned in any earlier round $1, \ldots, k-1$.

Boston Mechanism

The Boston mechanism works as follows. Each agent submits a strict ordinal ranking over all of the objects. The mechanism draws an ordering of the agents randomly from the uniform distribution over all agent orderings, which will be used as a "tie-breaker" below. The mechanism then proceeds in rounds as follows:

- Round 1: The mechanism considers the top-ranked object of each agent. If only one agent has ranked an object first, the object is assigned to that agent. If more than one agent has ranked an object first, the mechanism assigns the object to the agent among them who was ranked highest in the random agent ordering drawn above. The agents and objects that were assigned leave the market. If all agents have been assigned an object, the mechanism ends. Otherwise, all agents and objects that were not assigned in this round proceed to the next round.

- Round 2: Only agents who were not assigned an object in the first round participate. The mechanism looks at the second-ranked choices of all such agents. If only one agent has ranked an object second, the object is assigned to that agent. If more than one agent has ranked an object second, the mechanism assigns the object to the agent among them who was ranked highest in the random agent ordering drawn above. The agents and objects that were assigned leave the market. If all agents have been assigned an object, the mechanism ends. Otherwise, all agents and objects that were not assigned in this round proceed to the next round.

- Round $k = 3, \ldots, M$: The mechanism proceeds exactly as in round 2, except it now considers the k^{th}-ranked objects in the rank lists of the agents who participate in this round.

The definitions of RSD and Boston given above are simplifications of a more general class of mechanisms. It is easy to extend the mechanisms to incorporate features such as multiple copies of each object, outside options, and priority lists over the agents for each object, all of which are common in settings such as school choice (see, for instance, Abdulkadiroğlu and Sönmez (2003)). We use these definitions because they are sufficient to capture the properties of assignment mechanisms that are the focus of our investigation while still remaining parsimonious enough to design a lab experiment to cleanly test our hypotheses of rankings-dependent preferences.

2.2 Incentives and Welfare

Two key concerns when designing any allocation mechanism are the incentives they provide participants with regard to reporting their preferences and the equilibrium welfare of the resulting allocation. In this section, we briefly discuss these issues for both RSD and Boston. After discussing the details of the experimental design, we will return to these issues with formal theorems and experimental hypotheses based on them.

Truth-telling

A common desideratum when designing a matching mechanism is strategyproofness, the property that it is a weakly dominant strategy in the mechanism-induced game for each agent to submit their true preference ranking to the mechanism. In the standard matching literature where behavior is not rankings-dependent (i.e., $p(j) = 0$ for all j), this corresponds to submitting the preference ranking that lists objects in decreasing order of fundamental value. This strategy is, for obvious reasons, called the *truthful strategy.* Without rankings-dependent utility, RSD is well-known to be strategyproof while Boston is not. Intuitively, in Boston an agent with fundamental values $v_i(x) > v_i(y)$ may want to lie and report yP_ix, if x is likely to be very popular while y is almost as good and easier to get, because failing to get x might lead to getting an even worse object z. Indeed, Troyan and Morrill (2020) show that not only is Boston manipulable, it is *obviously* manipulable, and there is empirical evidence from school choice data that show families engaging in precisely this type of non-truth-telling behavior (Dur et al., 2018).

With the introduction of rankings-dependent utility in Equation (1), the strategyproofness property is not well-defined because there is no one-to-one mapping of utility to a preference ranking of the objects. Nevertheless, it is straightforward to still define the truthful strategy with respect to fundamental values exactly as above as the strategy that submits a preference ranking that lists objects in decreasing order of fundamental value. Though this terminology is not as descriptive in this model where fundamental value does not reflect utility, we continue to use it to maintain consistency with the previous literature. Due to the rankings-dependent utility, it is no longer the case that choosing the truthful strategy is weakly dominant in RSD. There is extensive empirical and experimental evidence (see the first paragraph in the introduction for references) that agents do not submit the truthful strategy to the mechanism in deferred acceptance, a mechanism that is equivalent to RSD in our setup. Indeed, this is the main motivation for incorporating and investigating rankings-dependent utility in the RSD mechanism.

Efficiency and Welfare

Without rankings-dependent utility, RSD always produces an assignment that is Pareto efficient ex-post: the first agent in the randomly chosen RSD ordering, say i_1, is assigned to her top-ranked object. The second agent, say i_2, is assigned to her highest-ranked object that was not taken by i_1, so the only way to make i_2 better off is to give her the object that went to i_1 (assuming i_2 prefers it), but this makes i_1 worse off. Continuing this argument, it is easy to see that each successive agent can only be made better off by taking the assignment of some earlier agent, and thus at least one of these agents must be worse off.

However, once agents' fundamental values and rankings-dependent utilities are taken into account (as opposed to just ordinal preferences), RSD may no longer perform as well on welfare grounds. For instance, consider 3 agents and 3 goods, with the following fundamental values.

	$v_i(x)$	$v_i(y)$	$v_i(z)$
i_1	1	1	1
i_2	0.7	0.7	0.7
i_3	0	0	0

Also, let $\rho(1) = 0.1$ for all agents, and $\rho(2) = \rho(3) = 0$. It is easy to check that truth-telling with respect to fundamental values (i.e., all agents report ordinal preferences $xPyPz$) is the unique equilibrium of RSD so the issue in the previous section is not relevant. No matter the random ordering, all final allocations result in a sum of utilities that is equal to $W^{SD} = (1 + 0.1) + 0.7 + 0 = 1.9$.

Next, consider Boston with the following strategies: agents i_1 and i_2 report truthfully with respect to fundamental values, while agent i_3 reports $yPzPx$. It can be checked that this is an equilibrium. In this equilibrium, one of the agents—agent i_3 in our example—reports her second-best good, good y, first. This ensures that she gets y for sure, but also leaves the remaining two agents with a higher chance of receiving the best good, x (in particular, agents i_1 and i_2 each have a 50/50 chance of receiving x and z). Since the preferences of i_1 and i_2 are symmetric, any final allocation results in a sum of utilities $W^B = (1 + 0.1) + (0.7 + 0.1) + 0 = 2 > W^{SD}$.

Thus, Boston results in a greater overall total welfare. Intuitively, the reason is that Boston incentivizes some agents to rank the good with the second-highest fundamental value first. In equilibrium, this means that there will be two agents who are getting a good ranked first, and so, both of these agents will receive the rankings-dependent utility $\rho(1)$. This is our motivation for comparing the RSD and Boston mechanisms in our experiment. In the next section, we will provide formal theorems and corresponding experimental hypotheses that generalize these intuitions for our experimental environment.[6]

3 Experimental Design

The experiment consists of two phases that participants complete in succession. We have two treatments that only differ in the matching mechanism run in Phase II. These treatments are *RSD* which runs the random serial dictatorship and *Boston* which runs the Boston mechanism. Hereafter, we will continue to use italics for the treatment names and regular text for the mechanisms. We will describe each phase here, and the full instructions are provided in Appendix D.

3.1 The Two Phase Experiment

Phase I

Phase I is identical for both treatments. The participants complete 22 incentivized tasks. The first 20 tasks consist of valuing 20 different objects with the multiple price list elicitation

[6]There is a recent strand of literature that also emphasizes that non-strategyproof mechanisms may outperform strategyproof ones in equilibrium even with just standard preferences because the opportunity to "misrepresent" their preferences gives a channel by which agents can express some information on their cardinal utilities (Abdulkadiroğlu et al., 2011; Troyan, 2012; Abdulkadiroğlu et al., 2015; Fragiadakis and Troyan, 2019). While this channel is still present in our model, additionally including rankings-dependent utility will amplify the welfare gains of non-strategyproof mechanisms.

method. See Appendix C for a description of each of the 20 objects including the monetary values (measured as the prices on Amazon.com where we purchased the objects) and average elicited values.

The multiple price lists are framed as willingness to accept (WTA) and consist of two screens. The first screen elicits the value of each object in dollar increments and the second screen elicits the value in two cent increments. In particular, for each of the 20 objects, the participants are told they have been allocated the object and then have the opportunity to exchange the object for various amounts of money. On the first screen, they see a list where each row represents keeping the object or exchanging it for an amount of money that ranges from \$1.00 to \$50.00. The participants choose the row that is the last row where they prefer keeping the object to exchanging it. We provide a screen shot in Figure 2a in Appendix D where the participant has selected the row with the dollar amount \$16.00. The participants can change their minds and click a different row, and then confirm their choices once they have come to a final decision. The second screen is presented identically to the first except that the amounts of money range from \$$x$.02 to \$$x$+1 in \$0.02 increments where x is the value of the last row (the one they had selected) from the first screen. We provide a screen shot in Figure 2b in Appendix D where the participant has selected the dollar amount \$16.56.

In order to assist the participants in this valuation task, pictures of each object are provided on the screens as the objects are valued. We also had the physical objects at the front of the room and would bring them over for further inspection on request. All participants value the frisbee first, the set of picture frames second, the cable spirals third, and the final 17 objects in random order. The reason to fix the first three objects is to allow participants to gain familiarity with the elicitation method. These three objects are not relevant for Phase II and are not analyzed in the results section.

If one of these valuation tasks is selected for payment for a given participant, we pay the participant in the standard way for multiple price lists. We randomly and uniformly draw a number between 1 and 50 which corresponds to a dollar amount the participant will receive. The participant then receives the object if they had indicated that they preferred the object to that amount of money and the dollar amount if they had indicated that they preferred the amount of money to the object. In the case that the randomly drawn number is exactly equal to the dollar amount in the last row where they would keep the object (the row they had clicked on the first screen), then a second randomly and uniformly drawn number between 1 and 50 is generated. In this case, the number corresponds to the two cent increments on the second screen (1 was \$0.02, 2 was \$0.04, etc.) and the participant receives the object if they had indicated that they preferred the object to the amount of money and the dollar and cents if they had indicated the opposite.

The final two tasks are incentivized risk aversion and loss aversion elicitation tasks. The risk aversion task is the classic Holt and Laury (2002) price list task. To be consistent with the previous 20 tasks, there are 50 rows of lottery choices between Lottery A, high payoff of \$24.00 with $x\%$ chance and low payoff of \$20.00 with $100 - x\%$ chance, and Lottery B, high payoff of \$38.00 with $x\%$ chance and low payoff of \$12.00 with $100 - x\%$ chance. The chance of the high payoff, x, increases across rows from 2% to 100%. The participants select the last row where they preferred Lottery A. If this task is selected for payment for a given participant, we randomly and uniformly draw a number between 1 and 50 and run Lottery A if that number is less than or equal to the number of the last row where they preferred Lottery A, and we run Lottery B otherwise.

The loss aversion task compares risky choices with gains to risky choices with losses. Again, there are 50 rows of lottery choices. For this task, the participant chooses between Option A, \$20.00 for sure plus a 50% chance of a \$10.00 bonus, and Option B, \$30.00 for sure plus a 50% chance of losing \$x. The loss in Option B varies from \$20.00 to \$0 in increments of \$0.40 across the rows. The participants select the last row where they prefer Option A. If this task is selected for payment for a given participant, we randomly and uniformly draw a number between 1 and 50 and run Option A if that number is less than or equal to the number of the last row where they preferred Option A, and we run Option B otherwise.

If either of the above tasks is selected for payment, we again pay the participant in the standard way by drawing a random number uniformly from 1 to 50 and implementing the choice of the participant (running the selected lottery) in the row corresponding to the random draw. Finally, the participants complete an unincentivized questionnaire to conclude the phase. The questionnaire includes the standard cognitive reflection task questions as well as demographics and academic endeavors.

Phase II

After completing the 22 tasks and survey in Phase I, the participants move on to Phase II. In this phase, the participants are randomly matched into groups of five to engage in a matching market with the five goods. We use the same five goods for Phase II in every session, and they are all taken from the set of twenty goods valued in Phase I. The goods for Phase II are:

1. Fjallraven backpack: A small 16L gray backpack from the popular brand Fjallraven with a monetary value of \$66.95.[7]

2. Hydroflask waterbottle: A light-blue reusable water bottle from the popular brand Hydroflask with a monetary value of \$32.96.

3. Moleskine notebook: A notebook with 192 pages and a high-quality black cover from the popular brand Moleskine with a monetary value of \$21.90.

4. Blue ceramic mug: A generic blue ceramic mug with a monetary value of \$11.99.

5. Set of 4 Uni-ball pens: A set of four fine-point black rollerball pens from the well-known brand Uni-ball with a monetary value of \$6.88.

Each participant engaged in a matching market corresponding to their treatment, *RSD* or *Boston*, just one time. They then value the good they received from the mechanism using the exact same multiple price list elicitation method used for the 20 goods valued in Phase I.

Remark 1. The participants only engage in a single matching market. We chose this method so that they would feel greater ownership of the good they received. In order to make sure the

[7] The goods were purchased on Amazon.com. The prices on Amazon vary slightly from day to day so the monetary values given here are approximate.

participants fully understood the procedure we presented them with instructions that included an example (with different goods than in the actual market), had them complete a quiz regarding a second example which required them to answer all the questions correctly before moving on, and provided them with an 8 minute practice period in which they could engage in as many markets as they wanted against robot players. We feel comfortable that they understood the mechanisms, because they almost all received 100% on the quiz the first time, rarely used the full 8 minutes to practice against the robots, and because neither the Boston nor the RSD mechanisms are particularly complicated mechanisms. All the time preparing them for the matching mechanism had an added bonus: there was a lot of time between valuing the good they received in Phase II and valuing that good in Phase I so we can be confident that most participants forgot their Phase I valuation. This is also the reason why we had participants value 20 objects, 15 of which were not relevant for the experiment, and why we put the questionnaire at the end of Phase I rather than at the end of the experiment.

3.2 Theoretical Predictions and Experimental Hypotheses

Rankings-Dependent Utility

We will analyze four main behavioral hypotheses in the results section. Our first and main hypothesis assesses whether preferences are indeed rankings-dependent.

Recall our model with rankings-dependent utility, Equation (1):

$$u_i(x, P_i) = v_i(x) + \rho(j).$$

To test the hypothesis that preferences are rankings-dependent, we are interested in the term $\rho(j)$. Our experimental design allows us to recover exactly this. Suppose that, in the matching mechanism in Phase II, a participant i receives object x, which was reported as her j^{th} ranked good. The valuation elicitation at the end of Phase II measures $v_i(x) + \rho(j)$. In Phase I, the participant also valued object x, independent of any mechanism or ranking context. Thus, the Phase I valuation of object x is $v_i(x)$. So the net value, $NV(j)$, the difference in valuations for the same object x between Phase II and Phase I, is the rankings-dependent component $\rho(j)$:

$$NV(j) = \text{Phase I value - Phase II value} = (v_i(x) + \rho(j)) - v_i(x) = \rho(j)$$

If $\rho(j)$ is decreasing in j as assumed in Section 3, the same should be true for the net value between Phase I and Phase II. This gives rise to our first hypothesis.

Hypothesis 1. *Participants prefer to get objects they rank higher, and so $NV(j)$ will be decreasing in submitted rank j.*

Truth-telling

Our next two hypotheses concern truth-telling and preference-reporting. Recall that, in the standard matching model without rankings-dependent utility, RSD is strategyproof while the Boston mechanism is not. In the case of Boston, agents may manipulate their preferences implied by their fundamental values by ranking popular objects lower, because they are likely to be taken in earlier rounds of the mechanism. But how do these properties extend to the case of rankings-dependent utility?

Truth-telling with respect to fundamental values will no longer be a dominant strategy of either mechanism, because agents may want to manipulate and rank objects that are less popular highly so that they receive a higher-ranked object and the corresponding increase in $\rho(j)$. Intuitively, there are two reasons an agent might want to deviating from truth-telling in Boston: (1) to avoid losing out on a middle-ranked object if she is unassigned in round one and (2) to obtain a higher $\rho(j)$. For SD, only (2) matters, and so we expect higher rates of truth-telling in SD compared to Boston.

Before stating our hypotheses, we will formalize the intuition from the previous paragraph. To state our formal theorems, we make the following assumption on preferences:

Assumption: For all $x_j \in X$, $v_i(x_j) = v_{i'}(x_j) := v_j$ for all $i, i' \in I$ and that these values are common knowledge. Further, $v_1, v_2 \gg \bar{v} = v_3 = \cdots = v_n$.

In words, we assume that there is complete information, a common ordinal preference, and that objects x_1 and x_2 are much better than the other objects.[8] The motivation for this assumption is that goods 1 and 2 are strongly enough preferred by all agents to all other goods, so that the main strategic decision is whether to rank x_1 or x_2 first. The exact goods in the experiment were chosen with this model of valuations in mind: given the popularity of Fjallraven backpacks and Hydroflask water bottles to our participant pool of undergraduate students, we expected these two objects to be the most popular, and strongly preferred to the other three objects (a notebook, coffee mug, and pens). Indeed, 84% of the participants value the backpack and water bottle as the two best goods, as measured by the Phase I elicitation. Table C in the appendix shows that the average elicited values for the five objects used in the matching mechanism were \$28.24 (backpack), \$22.56 (water bottle), \$9.11 (notebook), \$6.53 (mug) and \$5.33 (pens).

Given the preference assumption, for the theory, we focus on equilibria that have the following structure. For RSD: n_1 of the agents rank x_1 first and x_2 second; the remaining $n - n_1$ agents rank x_2 first and x_1 second; for the remaining goods $x_3, \ldots, x_n$, each agent draws a ranking of these goods uniformly at random from all possible rankings. For Boston: n_1 agents rank x_1 first; the remaining $n - n_1$ agents rank x_2 first; for the remaining goods $x_3, \ldots, x_n$, as in RSD, each agent draws a ranking of these goods uniformly at random from all possible rankings. They rank these goods immediately below their top choice, and rank the remaining good that was not their top choice (either x_1 or x_2) at the bottom of their list.

In other words, in RSD, all agents rank x_1 and x_2 first and second, in some order. In Boston, each agent ranks either x_1 or x_2 first, and puts the other option not chosen last in their preferences. This is in line with our motivation that the main source of competition is over goods x_1 and x_2. Thus, an agent's strategy boils down to whether to rank x_1 or x_2 first, and the key equilibrium object to solve for is the exact number of agents that choose to rank

[8] Though unlikely to ever hold exactly, common ordinal preferences is an assumption commonly made to approximate highly correlated preferences while still maintaining analytic tractability (see, e.g., Abdulkadiroğlu et al. (2011), Featherstone and Niederle (2016), and Fragiadakis and Troyan (2019)). As we explain in this paragraph, we think our preference assumption captures well the fundamental strategic tension we wanted to induce with our selection of objects, which is whether to rank the backpack or the water bottle first.

x_1 (the best object according to fundamental value) first in their reported preferences in RSD versus Boston. As we show in the proof of Theorem 2, this number is uniquely determined for each mechanism.[9]

Theorem 1. *Suppose the preference assumption holds. If truth-telling with respect to the fundamental value is an equilibrium of the Boston mechanism, truth-telling is also an equilibrium of RSD.*

This theorem suggests that there should be more truth-telling in RSD compared to Boston. Intuitively, truth-telling is an equilibrium of the Boston mechanism when the fundamental value v_1 is so high that it is worth it for every agent to enter the round 1 lottery for x_1, rather than deviating by ranking x_2 first and receiving payoff $v_2 + \rho(1)$ for sure. In RSD, this deviation is even less profitable, because it does not guarantee x_2 with certainty (though it does guarantee the deviating agent will not get x_1). Thus, if the deviation is not profitable in Boston, it will not be profitable in RSD, either, and so truth-telling with respect to fundamental values is an equilibrium of both mechanisms.

Our experimental design allows us to measure each participant's rankings with respect to fundamental value, because the fundamental values are elicited in Phase I. This gives our second testable hypothesis.

Hypothesis 2. *Participants will submit preferences in Phase II that are truthful with respect to fundamental values as measured by the elicitation in Phase I more in RSD than in Boston.*

Theorem 1 only discusses the truth-telling equilibrium. It may of course be that truth-telling is not an equilibrium of either Boston or RSD: Once rankings-dependent utility is introduced, even in RSD, an agent may want to deviate from truthful reporting in order to obtain the additional rankings-dependent utility term $\rho(1)$. Even in this case, however, we still expect more agents to misreport by ranking x_2 first in the Boston mechanism, because of the "double incentives" to do so (see (1) and (2) above). This is formalized in the next theorem.

Theorem 2. *Suppose the preference assumption holds. Then, in any equilibrium of RSD, weakly more agents rank good x_1 first in their preference list than in any equilibrium of Boston.*

Notice that the formal theorem only says that weakly more agents will rank x_1 first in RSD. It is possible for some values that the number of such agents is the same, but there also exist values for which the comparison is a strict inequality (i.e., strictly more agents do not follow the truthful strategy in Boston). This suggests our third hypothesis.

Hypothesis 3. *In Phase II, more participants will rank the good with the highest elicited Phase I value first in their preference list in RSD than in Boston.*

[9] There is one caveat to the strategies defined above, which is that if all agents rank x_1 first in Boston ($n_1 = n$), then, given the preference assumption, it is optimal for everyone to rank x_2 second, rather than last, in their preference list. Essentially, this happens when $v_1 \gg v_2$ to the extent that it is worth it for every agent to enter the round 1 lottery for x_1, in which case x_2 will still be available in round 2 of the mechanism, and Boston effectively becomes equivalent to RSD. The theorems stated below still hold regardless. See the proofs in Appendix A for details.

Welfare

Our final hypothesis concerns the overall welfare of the two mechanisms. Arguably, this is the most important feature of any mechanism, because the ultimate goal of any allocation mechanism is to produce an outcome that maximizes participant satisfaction, taking into account all components of utility, including fundamental values and rankings-dependent components.

Theorem 3. *Suppose the preference assumption holds. The equilibrium total welfare of the Boston mechanism is weakly higher than the equilibrium total welfare of RSD.*

The intuition for this result is that, given our preference assumption, the sum of the fundamental value components of total welfare is the same for any allocation.[10] Thus, the welfare comparison between any two mechanisms is determined by the sum of the rankings-dependent utility terms:

$$W = \sum_{j=1}^{n} (\#\text{ agents who receive their } j^{th} \text{ ranked good}) \times \rho(j)$$

In the Boston mechanism, in equilibrium, at least one agent ranks good x_2 first and receives it for sure, and so there are two agents who receive rankings-dependent utility $\rho(1)$.[11] In RSD, even if some agents do rank x_2 first, there is still a non-zero probability that the good will go to an agent who ranked it second, and who thus will receiving rankings-dependent utility $\rho(2) < \rho(1)$, resulting in lower total welfare for RSD. The full details of the argument can be found in the appendix.

In the experiment, we measure the total welfare of a mechanism as the sum of the elicited values for the mechanism allocation in Phase II. This gives our final hypothesis.

Hypothesis 4. *The sum of the elicited values in Phase II will be higher in Boston than in RSD.*

Remark 2. It is well-known that the elicitation method for values affects the elicited values. Most famously, the "endowment effect" which claims that participants value objects more when they own them than when they do not is very well-established (Ericson and Fuster, 2014). In our case, we frame the elicitation task as owning the objects in both phases so the endowment effect is netted out in the net value $NV(j)$. Hypotheses 2 through 4 are also comparisons made across values elicited the same way, and so any results inconsistent with the hypotheses can not be attributed to the choice of the multiple price list.

3.3 Procedures

All experiments were run at the University of Virginia VeconLab with undergraduate students recruited from the Darden BRAD lab recruitment pool. There were a total of 200 participants,

[10] As with Theorem 2, this theorem is stated as a weak inequality, but there exist values for which the inequality is strict.

[11] There is also the case that v_1 is so high that all agents rank x_1 first in Boston. However, in this case, all agents will also rank x_1 first in RSD (see Theorem 2) and the mechanisms become equivalent, so the welfare of the two mechanisms will be the same.

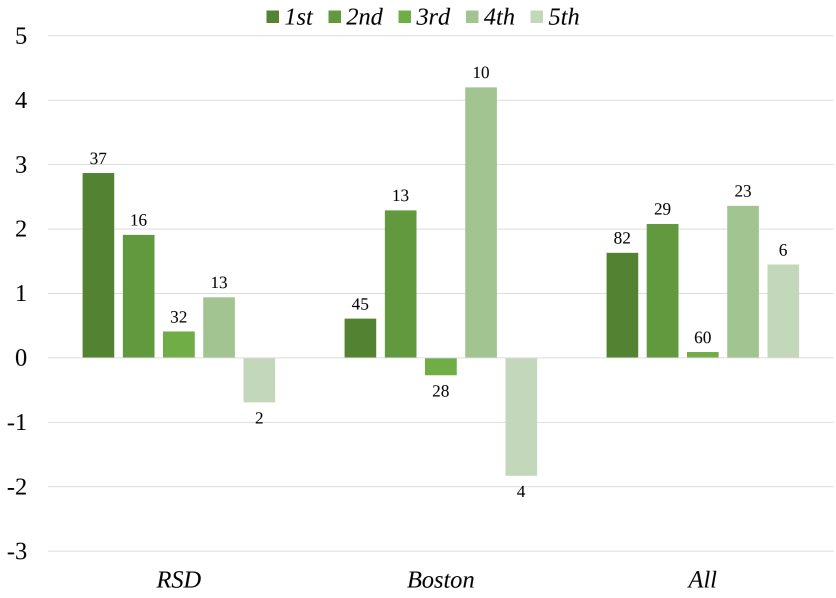


Figure 1: Net Values with # of obs. on ends of bars

100 for each of the two treatments in a between-subjects design. No feedback was provided before the end of the experiment so each participant is treated as an independent observation for the results. The experiment was programmed and run with z-Tree (Fischbacher, 2007).

One task was selected randomly for payment. There was a 50% chance it was the value elicitation task after the matching market and a 50% chance it was one of the 22 tasks in Phase I and, in the latter case, the task was randomly and uniformly selected. For the participants who earned money the average earnings were \$33.59. For the participants who received goods, the average monetary values of these goods was \$36.69.[12] Additionally, all participants received \$6 for showing up.

4 Results

In this section, we turn to the data to analyze the extent to which our experimental results provide support for our four hypotheses. For regressions we use stars to indicate significance at the usual levels ($*$ if $p < .1$, $**$ if $p < .05$, and $***$ if $p < .01$).

Hypothesis 1: Net Value

Recall that Hypothesis 1 argues that net value, the elicited value in Phase II minus the elicited value in Phase I, will be decreasing in submitted rank j. As a first look, Figure 1 provides the average net values for each rank for *RSD*, *Boston*, and both treatments.

For *RSD*, the results presented in Figure 1 are consistent with the hypothesis. Net value is

[12]In total, 144 participants earned money and 56 participants received goods.

almost monotonically decreasing, with the exception of the 3^{rd} and 4^{th} rankings (n is only 13 for the latter case, as most people received one of their top 3 ranked goods). Indeed, a non-parametric test rejects the null of no difference among the ranks in favor of the alternative that net value is decreasing in submitted rank (Jonckheere-Terpstra, $p = 0.019$). On the other hand Figure 1 seems to show no relationship between net value and submitted rank in *Boston*, and the non-parametric test does not reject the null in this case (Jonckheere-Terpstra, $p = 0.618$).

To explore these results in more detail, we turn to regression analysis. The regressions are presented in Table 1. For each treatment, we first regress net value on submitted rank alone (regressions (1) and (3)). We then add a number of potential influences on net value including Phase I elicited value (initial value), Holt-Laury switching point (risk aversion), the loss aversion task switching point (loss aversion), score out of three on the three cognitive reflection tasks (CRT score), a dummy for gender (female), the randomly selected order in which the good received in Phase II was valued in Phase I (Phase I order), the number of submitted preference orders in the practice session against robots (practice), and a dummy for whether the received good in Phase II was ranked higher than it would have been according to the implied fundamental values elicited in Phase I (truthful). These regressions are reported in columns (2) and (4), and the pooled regression with both treatments is presented in column (5).

Table 1: Net Value Regressions

Treatment	*RSD*		*Boston*		All
	(1)	(2)	(3)	(4)	(5)
rank	-0.8795*	-1.2358*	0.0979	-1.8063**	-1.5526***
initial value		-0.0570		-0.2768***	-0.1759***
risk aversion		0.0603		0.2676***	0.1679***
loss aversion		0.0015		-0.1331	-0.0769
CRT score		0.0222		-0.5717	-0.3484
female		-0.1220		-0.6279	-0.5844
Phase I order		0.1077		-0.1333	-0.0328
practice		-0.1485		-0.2194	-0.1477
truthful		-1.6938		-0.7317	-1.1834
_cons	3.6033***	4.5190	0.6334	10.4651**	8.1221**
No. of Obs.	100	100	100	100	200
R-Squared	0.03	0.11	0.00	0.27	0.16

In line with the non-parametric tests, Table 1 provides support for the hypothesis that net value is decreasing in rank of received good for *RSD*. Net value decreases by \$0.88 or \$1.24 per ranking depending on whether controls are included or not, and the estimates are marginally significant. Surprisingly, in contrast to the non-parametric tests, Table 1 also provides support for the hypothesis for *Boston* when controls are included. For *Boston*, net value decreases by \$1.81 per ranking when controls are included and the estimate is significant. Pooling the data, there is a strongly significant decline of \$1.55 per rank.

Additionally, in *Boston*, participants' initial values negatively impact net value while risk aversion positively impacts net value. The effect from initial value must be carefully evaluated, because initial value is part of the net value. Observe that the estimate -0.2768 means that an increase of \$1.00 in initial value decreases net value by \$0.28, but this can also be translated and interpreted as an increase of \$1.00 in initial value increases final value (the Phase II valuation) by \$0.72. Combined with the positive intercept in the regression, the estimates imply that small initial values are increased and large initial values are decreased. One interpretation of the result then is reversion to the mean; participants perhaps value objects with noise and so low Phase I valuations are more likely to be increased while high Phase I valuations are more likely to be decreased.[13]

The effect of risk aversion is harder to understand. As we will discuss in the next section, risk aversion impacts truth-telling in *Boston*. Briefly, we show there that risk averse participants are more likely to highly rank a good that they do not like as much but is less popular. Here, this means that when they get this less popular good their net values increase. We are not as confident as to why this may be the case, but one possibility is utility from the relief that they do indeed obtain the good they moved up in their rankings.

Given the discrepancy between the raw data and the regression analysis (particularly with the difficult to interpret positive effect of risk aversion in *Boston*), we think it is prudent to dive a little deeper. As a check of robustness, we ran additional regressions with dummy variables for final rank and for the good received. The first set of dummies allows for a non-linear relationship between rank and net value, while the second allows for the possibility that there is something inherent about the good itself. In order to not have too many regressors relative to data, we only included the controls for initial value and risk aversion. The results are presented in Table 5 in the appendix.

These results are show some similarities and some differences to Table 1. Most notably, the support for the hypothesis goes away. In particular, in *RSD*, the estimates that are significant are for receiving the highest-monetary goods, the backpack and the waterbottle.[14] The results for *Boston* indicate it is only initial value and risk aversion that matter. Putting them together, there is no evidence in support of the hypothesis in Table 5. So is Table 1 or Table 5 correct? Table 5 has higher values of R^2, but we also do not want to say it is more correct than Table 1. Indeed, the reason we have it as a check of robustness is that there are two reasons to be wary of the results in Table 5. First, the dummy variables reduce the power of the estimates for the effects of rank (essentially only considering 20 data points at a time rather than the whole set of 100). Indeed the estimates for ranks in *Boston* are negative indicating that rank does decrease net value, just not significantly, which could be an issue of power. Second, there is clearly high collinearity between the goods received, the initial value, and the final ranks. For example, Table 1 indicates that getting your 1^{st} ranked good is best for net value while Table 5 indicates that getting the backpack is best for net value. Of course, these two outcomes are highly correlated (and backpacks also had the largest initial values so that is correlated as well).

[13]For example, a small initial value is more likely to come from a participant who under-valued the good in Phase I relative to their true fundamental value and is therefore more likely to increase the valuation in Phase II.

[14]The mug is omitted. So the coefficients are interpreted with respect to the mug which has the smallest net value.

This is the reason we omit the good dummies in our main regressions reported in Table 1. So, to conclude, we do find evidence in favor of Hypothesis 1, particularly for *RSD*, but we admit there is some mitigating evidence as well.

Hypotheses 2 and 3: Truth-telling

We now move to assessing truth-telling (with respect to fundamental value) and Hypotheses 2 and 3. To assess Hypothesis 2, we begin with reporting the proportion of participants whose submitted rankings list in Phase II is truthful with respect to the fundamental values elicited in Phase I.[15]

When testing Hypothesis 2, we classify an agent as truth-telling only if all five goods were ranked in the same order as the fundamental value elicited in Phase I. However, this could be an overly restrictive definition of truth-telling, because the goods were intentionally chosen such that two of them (the backpack and water bottle) were likely to be much better than the other three to most people. To the extent that the remaining three goods (notebook, mug, pen) were viewed as significantly worse than the top two, and similar in value amongst themselves, participants may have been more focused on the choice between the top two goods and submitted "noisy" preferences over the remaining three goods. There also is likely to be noise in the Phase I elicitation procedure itself, and since the notebook, mug, and pen were similar (and lower) in value, this could be a cause of apparently non-truthful reporting over all 5 goods in Phase II.[16]

We deal with this issue in two ways. First, we also assess Hypothesis 3, which looks at the number of agents who report their top good truthfully. Second, we consider submitted rankings list in Phase II that are truthful with respect to the fundamental values elicited in Phase I with some noise. In particular, we chose an ad-hoc cutoff of $2 and classify preferences as truthful as long as they do not rank a good in Phase II higher than another good with an elicited value that is $2 more. All the proportions are presented in Table 2.

Table 2: Truth-telling Rates

Measure	*RSD*	*Boston*	All
Exact			
All	0.40	0.37	0.39
Top Choice	0.78	0.79	0.79
Top 2 Choices	0.60	0.60	0.60
Up to $2 differences			
All	0.58	0.53	0.56
Top Choice	0.82	0.83	0.83
Top 2 Choices	0.70	0.67	0.69

Of course, as the leniency allowed in truthful reporting increases, the proportions of truthful reporting increases. But the main result, regardless of how truth-telling is measured, is that

[15] The preferences elicited in Phase I are not always strict, in which case listing the indifferent goods in any order is classified as truthful.

[16] Indeed, the interpretation of the initial value coefficient in Table 1 suggests that this is the case.

there are no differences in truthful reporting between the two treatments. The non-parametric Wilcoxon ranksum test confirms this impression for all 6 pairwise comparisons between the treatments. We find no support that Hypotheses 2 and 3 hold strictly. We also find it interesting that even in the most lenient case, about 20% of participants are putting a good at the top of their submitted list that they valued more than $2 less than their top-value good.

To provide a clearer picture of what impacts truth-telling, we ran probit regressions with a dependent variable equal to 1 for participants who chose the truth and the independent variables risk aversion, loss aversion, CRT score, female, and practice (all measured the same as in the net value regressions). We report the results for the case where truth-telling is measured as the top 2 choices in the main text, because we think that misstating the bottom 3 goods is mostly noise and it is exactly these two goods for which there is an incentive to misrepresent one's preferences in the Boston mechanism. The other two measures are presented in Tables 6 and 7 in the Appendix where we find estimate values are quite similar although the levels of significance vary somewhat.

Table 3: Truth-telling Top 2 Choices Regressions

Treatment	*RSD*		*Boston*		All	
	(1)	(2)	(3)	(4)	(5)	(6)
risk aversion	-0.0160	-0.0117	-0.0255*	-0.0196	-0.0183*	-0.0147
loss aversion	0.0051	-0.0053	-0.0060	-0.0013	-0.0004	-0.0029
CRT score	0.2477*	0.2093	0.1546	0.0880	0.2011**	0.1544
female	0.5353	0.3808	0.5626**	0.4994*	0.5189**	0.4426**
practice	-0.0135	-0.0166	0.0369	0.0255	0.0094	0.0034
_cons	-0.1425	0.5057	0.3021	0.4284	0.0348	0.4137
Measure	Exact	$2 diff.	Exact	$2 diff.	Exact	$2 diff.
No. of Obs.	100	100	100	100	200	200

The regressions indicate that while truth-telling rates are similar across the two treatments, this may be due to different reasons. The estimate for CRT score is positive and marginally significant in *RSD* (with the exact measure, regression (1)) indicating that deeper-thinking participants tell the truth more often. So perhaps the non truth-telling in *RSD* is due to mistakes made by participants with low CRT scores rather than driven by rankings-dependent utility. The estimate for risk aversion is negative and marginally significant in *Boston* (with the exact measure, regression (3)) indicating that risk averse participants are less likely to tell the truth indicating that risk averse participants are less likely to tell the truth, which is line with a model of standard preferences. Finally, the estimate for female is positive and significant in *Boston* which is consistent with the finding that females are more averse to lying (Dreber and Johannesson, 2008; Erat and Gneezy, 2012).

Hypothesis 4: Welfare

Finally, we address Hypothesis 4 and welfare. Table 4 presents the results for welfare measured as the sum of participants elicited Phase II values for the goods they received in the matching mechanism.

Table 4: Welfare

RSD	*Boston*	All
90.29	89.97	90.13

As is clear in the table, there is no difference between the two treatments. This is confirmed with the Wilcoxon rank sum non-parametric test. In light of the lack of differences in truth-telling between the two treatments, it is not surprising that ultimately welfare does not differ as that was the basis on which we hypothesized the difference.

5 Conclusion

We investigate whether agent preferences over goods received via a matching market are influenced by how highly they ranked the object in their reported preference list. We design a laboratory experiment to test whether agents value a good more the higher that they rank it. A novel feature of our experiment is that we use real goods, rather than induced values. This more faithfully simulates real-world settings in which participants must form their own values over their potential outcomes, and allows us to directly measure for rankings-dependent utility by eliciting values for objects both inside and outside of the mechanism and taking the difference.

We find evidence to support the hypothesis that valuations are indeed influenced by reported rankings. While there are many reasons this could be the case (e.g., EBRD preferences), differentiating between all of the possible explanations is difficult, as these explanations will often be outcome-equivalent. Regardless of the reason, the existence of rankings-dependent utility has important potential implications. In particular, it may upend the standard welfare comparisons between mechanisms, with mechanisms that are designed to give more agents objects that they *report* as higher-ranked having the potential to increase welfare. While we find no such differences in welfare in our experimental setting, this could be due to the specific objects and mechanism details that were chosen. Given that rankings-dependent preferences do seem to exist in matching markets, it is natural to expect that this could differentially impact the welfare consequences of various mechanisms in other settings. All said, having established the existence of a rankings-dependent component to utility in this paper, more work needs to be done to explore more deeply its root causes as well as the potential welfare implications of this issue.

Finally, while we motivate the use of real goods as a way to directly measure rankings-dependence, we also think they are interesting to examine more generally. For example, are real goods the reason we find similar rates of truth-telling across a strategyproof and a non-strategyproof mechanism, and in a way similar to the no information treatments of Pais and Pintér (2008), or is it another feature of our environment? Without induced-value control treatments, we cannot say for sure, but we think answering this question, and other questions about the impacts of real vs. fictitious goods, is a promising and vastly under-studied area for future research in matching markets and game theory more generally.

A Proofs of Theorems

Proof of Theorem 1. By supposition of the theorem, assume that truth-telling is an equilibrium of the Boston mechanism. We will show that truth-telling is also an equilibrium of RSD by showing that if all $j \neq i$ report truthfully, then i's best response is also to report truthfully.

Let all agents other than $j \neq i$ report a preference ranking that respects the fundamental values, $P_j : x_1, \ldots, x_n$. Fixing these reports for $j \neq i$, let i's expected utility from any report P_i' be $EU^{SD}(P_i')$. If i also reports a ranking that respects the fundamental values, $P_i : x_1, \ldots, x_n$, this becomes

$$EU_i^{SD}(P_i) = \frac{1}{n}\sum_{j=1}^{n} u_i(x_j, P_i) = \frac{1}{n}\sum_{j=1}^{n}(v_j + \rho(j)). \tag{2}$$

Consider any alternative report for i, $P_i' \neq P_i$. Let k' be the lowest index such that object $x_{k'}$ is not ranked in the k'-th position, and let x_{k^*} be the object that is ranked k'-th, where by definition, $k^* > k'$. Notice that such a k^* exists, as otherwise $P_i' = P_i$. Under P_i', agent i will never receive any object $x_{k'}, \ldots, x_{k^*-1}$. To see this, let ℓ be i's (randomly drawn) order in the serial dictatorship. If $\ell < k'$, then i will receive object x_ℓ; if $k' \leq \ell \leq k^*$, object x_{k^*} is still available when it is i's turn to choose, and thus i receives object k^*; if $\ell > k^*$, then at i's turn, all objects $x_1, \ldots, x_{k^*}$ have been taken by earlier agents, and thus i receives some object from the set $x_{k^*+1}, \ldots, x_n$. This implies that i's utiliy from reporting P_i' is bounded above by

$$EU^{SD}(P_i') \leq \frac{1}{n}\sum_{j=1}^{k'-1}(v_j + \rho(j)) + \frac{n-k'+1}{n}(v_{k^*} + \rho(k')) \tag{3}$$

Now, consider the Boston mechanism, and the same report P_i' from above (continuing to assume that all other agents report truthfully in Boston). If i reports P_i' in the Boston mechanism, then her expected utility is

$$EU^{B}(P_i') = \frac{1}{n}\sum_{j=1}^{k'-1}(v_j + \rho(j)) + \frac{n-k'+1}{n}(v_{k^*}) + \rho(k')) \tag{4}$$

This is because, in Boston, if i is not assigned in one of the first $k'-1$ rounds, she is assigned to x_{k^*} in round k' with certainty. In this case, she has ranked this object k'-th, so her total utility is the term in parentheses at the end of equation (4). This event occurs with probability $(n - k' + 1)/n$.

Notice that if i reports truthfully with respect to fundamental values in Boston, her payoff is

$$EU_i^{B}(P_i) = \frac{1}{n}\sum_{j=1}^{n}(v_j + \rho(j)), \tag{5}$$

which is equivalent to the payoff from reporting P_i in SD; see equation (2). Finally, we have the following:

$$EU^{SD}(P_i') \leq EU^{B}(P_i') \leq EU^{B}(P_i) = EU^{SD}(P_i) \tag{6}$$

where the first inequality follows from equations (3) and (4), the second from the fact that truth-telling is assumed to be an equilibrium of Boston, and the last equality from equations (2) and (5). Equation (6) implies $EU^{SD}(P_i') \leq EU^{SD}(P_i)$, i.e., truth-telling is optimal in SD. ■

Proof of Theorem 2. First, consider Boston, and let n_1^B be the number of agents who rank x_1 first. It is trivial to see that $n_1^B = 0$ is never an equilibrium, because if no agent is ranking x_1 first, then any agent can deviate to a strategy that does so and receive payoff $v_1 + \rho(1)$ for sure, which is the highest possible attainable payoff.

Next, consider the case that $n_1^B = n$ is an equilibrium of the Boston mechanism, i.e., all agents rank x_1 first. In this case, each agent will also rank x_2 second (see footnote 9), and so receives x_1 with probability $1/n$ and x_2 with probability $1/n$, for an equilibrium utility of $(1/n) \times (v_1 + \rho(1)) + 1/n \times (v_2 + \rho(2)) + ((n-2)/n) \times (\bar{v} + \bar{\rho})$, where $\bar{\rho} = \frac{1}{n-2} \sum_{j=3}^{n} \rho(j)$. The last term, $\bar{v} + \bar{\rho}$, comes from the fact that for all agents not assigned in rounds 1 or 2, the assignment is just a random assignment of goods $x_3, \ldots, x_n$, and so by symmetry, any individual agent's payoff is just the average, $\bar{v} + \bar{\rho}$; this event happens with probability $(n-2)/n$. If any individual agent deviates to ranking x_2 first, she gets it for sure, with resulting utility $v_2 + \rho(1)$. Since this is an equilibrium, we conclude that $(1/n) \times (v_1 + \rho(1)) + 1/n \times (v_2 + \rho(2)) + ((n-2)/n) \times (\bar{v} + \bar{\rho}) \geq v_2 + \rho(1)$. Note also if $n_1^B = n$ is an equilibrium, then it is the unique equilibrium. This follows because if there were some other equilibrium in which $n_1^B < n$, then some agent i is not ranking x_1 first. Agent i will therefore never receive x_1, and so her payoff is bounded above by by $v_2 + \rho(1)$. If she deviates to ranking x_1 first and x_2 second, then her payoff is bounded below by $(1/n) \times (v_1 + \rho(1)) + 1/n \times (v_2 + \rho(2)) + ((n-2)/n) \times (\bar{v} + \bar{\rho})$. As was just shown, the latter is greater than the former, and so this deviation is profitable.

Now, consider SD. If all n agents rank x_1 first and x_2 second, then they once again receive payoff $(1/n) \times (v_1 + \rho(1)) + 1/n \times (v_2 + \rho(2)) + ((n-2)/n) \times (\bar{v} + \bar{\rho})$. If any agent deviates to ranking anything other than x_1 first,[17] they will never receive x_1, and so their payoff is bounded above by $v_2 + \rho(1)$, which, from the above calculation for Boston, is smaller than the expected payoff from ranking x_1 first and x_2 second. Thus, $n_1^{SD} = n$ is also an equilibrium of SD. Similar arguments as for the Boston case above show that this equilibrium is once again unique, and so $n_1^{SD} \geq n_1^B$, as required.

So, for the remainder of the proof, we restrict to $0 < n_1 < n$. Let $U^B(x_j, n_1)$ be the expected utility of an agent who ranks good x_j first when n_1 total agents (including this agent) rank x_1 first, and the remaining $n - n_1$ total agents rank x_2 first. Then, we can calculate:

$$U^B(x_1, n_1) = \frac{1}{n_1}(v_1 + \rho(1)) + \frac{n_1 - 1}{n_1}\delta$$
$$U^B(x_2, n_1) = \frac{1}{n - n_1}(v_2 + \rho(1)) + \frac{n - n_1 - 1}{n - n_1}\delta$$

where $\delta = \bar{v} + (\rho(2) + \rho(3) + \ldots + \rho(n-1))/(n-2)$.[18] The equations derive from the fact that

[17] This follows because if the agent is chosen first in the RSD ordering, she receives her first-ranked object which is different from x_1, while if she is not chosen first, then whoever is chosen first takes x_1.

[18] This is slightly different than $\bar{v} + \bar{\rho}$ calculated above for the $n_1 = n$ case, because here, agents begin ranking goods $x_3, \ldots, x_n$ second in their list, and so the summation of the rankings-dependent utility terms run from

if an agent ranks x_1 first in Boston, she has a $1/n_1$ chance of getting x_1. If she does not, good x_2 will not be available in round 2, and so she will get some good $x_3, \ldots, x_n$. There will always be exactly $n-2$ agents left after round 1 of the mechanism, and since all agents rank goods $x_3, \ldots, x_n$ the same, this becomes a random allocation of the remaining objects. The δ terms represent the total expected utility from this random allocation. Let

$$\Delta^B_{x_1 \to x_2}(n_1) = U^B(x_1, n_1) - U^B(x_2, n_1 - 1)$$

be the change in utility for an agent whose equilibrium strategy is to rank x_1 first, and who deviates to ranking x_2 first. Similarly, let

$$\Delta^B_{x_2 \to x_1}(n_1) = U^B(x_2, n_1) - U^B(x_1, n_1 + 1)$$

be the expected change for an agent whose equilibrium strategy is to rank x_2 first, but who deviates to ranking x_1 first.

For n_1^B to be an equilibrium, we need both $\Delta^B_{x_1 \to x_2}(n_1^B) \geq 0$ and $\Delta^B_{x_2 \to x_1}(n_1^B) \geq 0$, i.e., no agent has a profitable deviation. Algebra shows that these two equations reduce to

$$n_1^B \in \left[\frac{n(v_1 + \rho(1) - \delta)}{v_1 + v_2 + 2(\rho(1) - \delta)} - \frac{v_2 + \rho(1) - \delta}{v_1 + v_2 + 2(\rho(1) - \delta)}, \frac{n(v_1 + \rho(1) - \delta)}{v_1 + v_2 + 2(\rho(1) - \delta)} + \frac{v_1 + \rho(1) - \delta}{v_1 + v_2 + 2(\rho(1) - \delta)}\right]. \tag{7}$$

Subtracting the left endpoint from the right endpoint gives a range of length exactly 1. So, there will be a unique integer n_1^B in this range, which corresponds to the unique equilibrium number of agents who rank x_1 first in Boston.

Let n_1^{SD} be the equilibrium number of agents who rank x_1 first in SD. We will show that $n_1^{SD} \geq n_1^B$. For SD, the analogous equations to the above are:

$$U^{SD}(x_1, n_1) = \frac{1}{n}(v_1 + \rho(1)) + \frac{1}{n}\left(\frac{n_1 - 1}{n-1}(v_2 + \rho(2)) + \frac{n - n_1}{n-1}(v_1 + \rho(1))\right) + \frac{n-2}{n}\delta'$$
$$U^{SD}(x_2, n_1) = \frac{1}{n}(v_2 + \rho(1)) + \frac{1}{n}\left(\frac{n_1}{n-1}(v_2 + \rho(1)) + \frac{n - n_1 - 1}{n-1}(v_1 + \rho(2))\right) + \frac{n-2}{n}\delta'$$

In each equation above, there is a $1/n$ chance that the agent is ordered first, in which case she gets her first ranked good. If not, there is a $1/n$ chance she is ordered second. In the top equation, there is an $(n_1 - 1)/(n-1)$ chance the first agent was one of the remaining agents who ranked x_1 first, in which case agent i receives x_2, and a $n_2/(n-1)$ chance the first agent was one of the agents who ranked x_2 first, in which case i gets x_1 again. If i is ordered third or higher in the SD ordering, then both goods x_1 and x_2 are gone at her turn, and, as for the Boston case above, i's utility in this case is that from a random assignment of the remaining goods, represented by $\delta' = \bar{v} + \bar{\rho}$.[19]

Define $\Delta^{SD}_{x_1 \to x_2}(n_1)$ and $\Delta^{SD}_{x_2 \to x_1}(n_1)$ analogously to the above, except replacing SD for

$j = 2$ to $n-1$; they will never receive their n^{th} ranked object, which is the object among $\{x_1, x_2\}$ that they did not rank first.

[19]The term δ' here is slightly different from the δ term in the Boston equations above, because in RSD, agents rank x_1 and x_2 first and second, in some order, while in Boston, one of x_1 or x_2 is ranked last. However, for RSD, the δ' terms are less important, because they will cancel out when checking for profitable deviations below.

Boston. Similarly as for Boston, for n_1^{SD} to be an equilibrium, we need $\Delta^{SD}_{x_1\to x_2}(n_1^{SD}), \Delta^{SD}_{x_2\to x_1}(n_1^{SD}) \geq 0$. Algebra shows that these equations reduce to

$$n_1^{SD} \in \left[\alpha + \frac{1}{2}, \alpha - \frac{1}{2}\right] \tag{8}$$

where $\alpha = \frac{n}{2} + \frac{(n-1)}{2}\frac{v_1-v_2}{\rho(1)-\rho(2)}$. Once again, this range has total length 1, and so there is a unique equilibrium number of agents n_1^{SD} that rank x_1 first.

What remains to check is that $n_1^{SD} \geq n_1^B$. To show this, note first that the lower bounds in equations (7) and (8) are determined by the equations $\Delta^B_{x_2\to x_1}(n_1) \geq 0$ and $\Delta^{SD}_{x_2\to x_1}(n_1) \geq 0$ that ensure that those who rank x_2 first do not want to deviate to ranking x_1. Next, notice that

$$\frac{d\Delta^{SD}_{x_2\to x_1}(n_1)}{dn_1} = \frac{2(\rho(1)-\rho(2))}{n(n-1)} > 0,$$

which implies that $\Delta^{SD}_{x_2\to x_1}(n_1)$ is an increasing function. Now, evaluate $\Delta^{SD}_{x_2\to x_1}(n_1)$ at $\xi = \frac{n(v_1+\rho(1)-\delta)}{v_1+v_2+2(\rho(1)-\delta)} - \frac{v_2+\rho(1)-\delta}{v_1+v_2+2(\rho(1)-\delta)}$, which is the lower bound of equation (7):

$$\Delta^{SD}_{x_2\to x_1}(\xi) = -\frac{(v_1-v_2)((n-3)\rho(1)+(n+1)\rho(2)+(n-1)(v_1+v_2)-2(n-1)\delta)}{n(n-1)(v_1+v_2+2\rho(1)-2\delta)}$$

Now, since $\bar{v} < v_2$ and $\rho(j)$ is a decreasing function, we have

$$\delta = \bar{v} + \frac{1}{n-2}\sum_{j=2}^{n-1}\rho(j) < v_2 + \rho(2)$$

It is then simple to check that both the numerator and denominator of the above equation are positive, which implies that $\Delta^{SD}_{x_2\to x_1}(\xi) < 0$. Because the function is increasing, the crossover point that defines the lower bound of equation (8) must lie to the right of ξ. Thus, the range of equation (8) must be to the right of the range of equation (7) (they may overlap), which implies that $n_1^{SD} \geq n_1^B$. ∎

Proof of Theorem 3. The overall welfare from any mechanism is just the sum total of the utilities. Recall that an agent's utility is the sum of the fundamental value of the object she receives, v_j, and a rankings-dependent component: $u_i(x, P_i) = v_i(x) + \rho(j)$. Because $v_i(x_j) = v_j$ for all i, the sum of the fundamental values will be the same for any assignment. This means that the overall welfare of any mechanism ψ is determined by the sum of the rankings-dependent utility components:

$$W^{\psi} = \sum_{j=1}^{n}(\#\text{ agents who receive their } j^{th} \text{ ranked good}) \times \rho(j)$$

There are two cases. First, if $n_1^B = n$—that is, all agents choose to rank x_1 first in the Boston mechanism—then, by Theorem 2, $n_1^{SD} = n$ as well. The two mechanisms are then equivalent, and so $W^B = W^{SD}$.

Second, assume that $1 \leq n_1^B < n$ (recall that Theorem 2 shows that $n_1^B = 0$ is never an equilibrium of Boston). For Boston, we have

$$W^B = 2\rho(1) + \sum_{j=2}^{n-1} \rho(j).$$

The first term, $2\rho(1)$ comes from the fact that at least one agent is ranking x_1 first and at least one agent is ranking x_2 first, and so both of these agents are receiving their top-ranked object. Then, all remaining agents rank the remaining goods uniformly at random, and so, by symmetry, they are equally likely to get any of these goods. Notice that the summation starts at 2, because they place the good from $\{x_1, x_2\}$ that was not their top choice at the bottom of their rankings (and will never receive it).

For SD, the equation is

$$\begin{aligned} W^{SD} = \frac{n_1^{SD}}{n}\frac{n_1^{SD}-1}{n-1}(\rho(1)+\rho(2)) + 2 \times \frac{n_1^{SD}}{n}\frac{n-n_1^{SD}}{n-1}(\rho(1)+\rho(1)) \\ + \frac{n-n_1^{SD}}{n}\frac{n-n_1^{SD}-1}{n-1}(\rho(1)+\rho(2)) + \sum_{j=3}^{n}\rho(j) \end{aligned} \tag{9}$$

The first three terms derive from the probability that the first two agents are "$x_1 P_i x_2$" agents or "$x_2 P_i x_1$" agents. For instance, for the first term, there is an n_1^{SD}/n chance that the first agent in the order ranks $x_1 P_i x_2$ and, conditional on this, a $(n_1^{SD} - 1)/(n - 1)$ chance that the second agent has the same ranking. In this case, x_1 goes to an agent who ranked it first and x_2 goes to an agent who ranked it second, so the rankings-dependent component of welfare is $\rho(1) + \rho(2)$. The next two terms are calculated similarly, for all possible combinations of the preferences of the first two agents. Finally, the summation term at the end comes from the fact that all remaining agents rank $x_3, \ldots, x_n$ uniformly randomly, and so by symmetry, are equally likely to get any rank. Note that, unlike in Boston, the summation runs from $j = 3$ to n, because all agents rank x_1 and x_2 first in SD.

Now, notice that

$$\begin{aligned} W^{SD} &\leq \frac{n_1^{SD}}{n}\frac{n_1^{SD}-1}{n-1}(2\rho(1)) + 2 \times \frac{n_1^{SD}}{n}\frac{n-n_1^{SD}}{n-1}(2\rho(1)) + \frac{n-n_1^{SD}}{n}\frac{n-n_1^{SD}-1}{n-1}(2\rho(1)) + \sum_{j=3}^{n}\rho(j) \\ &= 2\rho(1)\left(\frac{n_1^{SD}}{n}\frac{n_1^{SD}-1}{n-1} + 2 \times \frac{n_1^{SD}}{n}\frac{n-n_1^{SD}}{n-1} + \frac{n-n_1^{SD}}{n}\frac{n-n_1^{SD}-1}{n-1}\right) + \sum_{j=3}^{n}\rho(j) \\ &= 2\rho(1) + \sum_{j=3}^{n}\rho(j) \\ &\leq W^B \end{aligned}$$

where the first line replaces $\rho(2)$ with $\rho(1) \geq \rho(2)$, the second factors out the $2\rho(1)$ terms, the third follows because the term in parentheses sums to 1, and the last follows from $\rho(j)$ being a decreasing function. ■

B Robustness Regressions

Table 5: Net Value Regressions with Dummies

Treatment	SD	Boston	All
	(1)	(2)	(3)
ranked 2nd	1.0655	-0.6298	1.2793
ranked 3rd	0.3480	-1.8964	-0.4721
ranked 4th	0.3046	-0.6471	0.4093
ranked 5th	-1.0711	-4.5459	-3.0656
backpack	10.3001***	5.4340	8.1776***
notebook	2.6014	-2.4099	-0.0261
waterbottle	7.5522***	4.2488	6.3798***
pens	1.7752	-4.0435*	-1.2110
initial value	-0.2130***	-0.4093***	-0.3186***
risk aversion	0.0654	0.1927**	0.1436***
_cons	-1.6011	2.3024	-0.3960
No. of Obs.	100	100	200
R-Squared	0.25	0.31	0.23

Table 6: Truth-telling All Choices Regressions

Treatment	RSD		Boston		All	
	(1)	(2)	(3)	(4)	(5)	(6)
risk aversion	-0.0224	-0.0142	-0.0211	-0.0270*	-0.0165	-0.0177*
loss aversion	0.0000	-0.0057	-0.0224	-0.0042	-0.0110	-0.0046
CRT score	0.1794	0.1956	0.2118	0.2475*	0.1863*	0.2250**
female	0.5718*	0.2012	0.4025	0.4182	0.4065*	0.3214
practice	-0.0701**	-0.0507*	0.0481*	0.0282	-0.0024	-0.0105
_cons	0.1589	0.6542	-0.0824	0.0730	-0.0909	0.2440
Measure	Exact	$2 diff.	Exact	$2 diff.	Exact	$2 diff.
No. of Obs.	100	100	100	100	200	200

Table 7: Truth-telling Top Choice Regressions

Treatment	RSD		Boston		All	
	(1)	(2)	(3)	(4)	(5)	(6)
risk aversion	-0.0054	-0.0087	-0.0310*	-0.0225	-0.0150	-0.0138
loss aversion	0.0076	0.0034	-0.0060	-0.0047	0.0004	-0.0008
CRT score	0.2041	0.1716	0.0538	0.0787	0.1151	0.1197
female	0.4395	0.3006	-0.0074	0.0962	0.1824	0.1757
practice	-0.0201	-0.0226	-0.0021	-0.0043	-0.0154	-0.0157
_cons	0.1928	0.7298	1.8629**	1.6134**	1.0138**	1.1603**
Measure	Exact	$2 diff.	Exact	$2 diff.	Exact	$2 diff.
No. of Obs.	100	100	100	100	200	200

C Phase I goods

Table 8: Phase I goods

Good	Description	Amazon Price	Avg. Elic. Val.
Backpack	**Fjallraven gray, 16L backpack**	**$66.95**	**$28.24**
Alarm clock	Aisuo bluetooth alarm/speaker/nightlight	$37.99	$19.46
Water bottle	**Hydroflask blue, 32 oz. water bottle**	**$32.96**	**$22.56**
Shower speaker	Donerton bluetooth waterproof speaker	$29.99	$18.75
Laptop stand	Ergonomic universal laptop stand	$26.99	$15.99
Outdoor blanket	Bearz blue waterproof blanket	$24.99	$14.04
Cold brewer	Takeya 1 qt. carafe	$21.00	$15.04
Picture frames	Set of 4 white, wooden frames	$20.99	$9.99
Popcorn set	3 bags of popcorn plus flavors	$22.00	$8.72
Notebook	**Moleskine 192 age, black notebook**	**$21.90**	**$9.11**
Tile	Bluetooth chip to track item on phone	$19.99	$14.98
Phone mount	Gooseneck phone holder with bracket	$19.79	$8.87
Popcorn popper	Silicone microwaveable popcorn maker	$14.99	$7.49
Charging pad	Anker phone-charging pad	$13.99	$14.57
Frisbee	Discraft yellow 175 g. Frisbee	$13.76	$6.60
Mug	**Blue ceramic mug**	**$11.99**	**$6.53**
Playing cards	Deck of black playing cards	$7.99	$4.93
Pens	**Set of 4 Uni-ball rollerball pens**	**$6.88**	**$5.33**
Keychain tool	Multi-tool that attaches to keys	$6.64	$5.11
Cable spirals	Set of 24 plastic, multicolor cable-protectors	$6.29	$5.03

Notes:

1. Phase II goods in bold.
2. Amazon prices taken from first purchase and do slightly change over time.

D Instructions and Screen Shots

Welcome and Phase I

Welcome. This is an experiment in decision making. Various research foundations and institutions have provided funding for this experiment and you will have the opportunity to make a considerable amount of money which will be paid to you at the end. Make sure you pay close attention to the instructions because the choices you make will influence the amount of money you will take home with you today. Please ask questions if any instructions are unclear.

Terminology: In the following instructions, we will say that the computer will make a **random** choice from a number of possibilities. This means that the computer will randomly select one of the possibilities with equal chance for each. If there are N possibilities, you can think of this as the computer rolling a die that has N sides, and choosing the possibility that comes up on the die.

This experiment will consist of two phases. We will hand out the instructions for each phase before you complete the phase. You will be paid for your choices in only one of the two phases which will be randomly selected by the computer. Your choices in each phase have no impact on later phases.

Your earnings may be an object or money, and at the end of the experiment, we will give you the object or the money to take home. You will see pictures of the object on your screen, but we have the actual objects here with us. If you want us to show you the actual object at any time, just raise your hand and we can bring it over. Everyone will also get $6 for participating.

Phase I

For this phase, you will be allocated an object that will be shown to you on your screen. Remember that we have the actual objects with us, so feel free to raise your hand if you would like us to bring one over for closer inspection.

The Task

After you are allocated your object, you will have the opportunity to give up the object in exchange for a certain amount of money. Whether you keep the object or receive money will be determined as follows:

First, you will see a screen with 50 rows. Each row is a choice between keeping the object or exchanging it for $1, $2, . . . , up to $50 . As we will explain carefully in the earnings section of the instructions below, if a row is selected for your earnings, you will take home the choice (object or money) that you selected in that row. Because it is time-consuming to have you click a button for every row, instead you only need to **click on the bullet for the object in the last row where you would keep the object over the specified number of dollars.** Because all rows above the one you select offer less money, we will fill in all of these rows with you selecting to keep the object. Similarly, because all rows below offer more money, we will fill in all of these rows with you selecting the money. After you click, you can change your mind by clicking on a different row and that row will become the last row where you keep the object. Click confirm when you have finalized your choice.

Next, you will see another screen with 50 rows. Each row is a choice between keeping the object or exchanging it for $x.02, $x.04, up to $x+1 where x is the dollar amount in the row you selected on the first screen. As for the first screen, you make just one choice. **Click on the bullet for the object in the last row where you would keep the object over the specified number of dollars.** As on the first screen, because all rows above have less money, we will automatically fill in all of these rows with you selecting the object over the money. Because all rows below contain more money, we will automatically fill in all these rows with you selecting the money over the object. After you click, you can change your mind by clicking on a different row and that row will become the last row where you keep the object. Click confirm when you have finalized your choice.

Procedures

You will do this task for **20 objects**. After you complete the task for the 20 objects, we will ask you to complete 2 additional tasks unrelated to this task for a total of 22 tasks in Phase I. The instructions for those tasks will be presented on the screen when you do them. Finally, we will have you answer a short questionnaire.

Earnings

If this phase is selected for earnings, the computer will first randomly select 1 of the 22 tasks and you will receive earnings for that task. If the selection is one of the 20 tasks described above, your earnings will be determined as follows. The computer will randomly select one of the 50 rows from the first screen. **We will then implement your choice from that row. That is, you will get the object to take home with you if you selected to keep the object in this row, or the amount of money if you selected the money.**

There is one exception: If the randomly selected row from the first screen is the last row where you would keep the object, the computer will randomly select one of the 50 rows from the second screen and **we will then implement your choice from that row.** Keep in mind that any row from any task could be selected so it is in your best interest to select the object when you prefer the object to the money and to select the money when you prefer the money to the object.

For the 2 additional tasks, how your earnings would be determined if the computer selected one of them will be explained to you when you complete those tasks .

Phase II: RSD

Phase II

For this phase, you will be divided into groups of 5 people, and for each group there will be 5 objects. Each member of the group will end up with exactly one of the objects. The object you get will be determined by your preferences for the objects, along with the preferences of your group members. Each group member will provide a **preference ranking** of the objects. Because there is only one of each object available, if multiple people list a particular object first, we will use a random tie-breaking procedure to determine who gets what. The details of how this is done are described below.

The Task

You will first be shown the 5 objects. Your task for this phase is to **submit a list ranking the objects from most-preferred to least-preferred.** After collecting everyone's rankings, the computer will randomly select a picking ordering of the 5 group members. The computer will then allocate an object to each group member as follows.

- The group member who is randomly chosen to go first will be allocated the first object on their submitted list.
- The group member who is randomly chosen to go second will be allocated the highest-ranked object on their list that was not taken by the first group member.
- The group member who is randomly chosen to go third will be allocated the highest-ranked object on their list that was not taken by the first or second group member.
- The group member who is randomly chosen to go fourth will be allocated the highest-ranked object on their list that was not taken by the first, second, or third group member.
- The group member who is randomly chosen to go fifth will be allocated the highest-ranked object on their list that was not taken by the first, second, third, or fourth group member.

It is up to you how to rank the objects. Given the rules of the procedure, there is no way to manipulate your rankings to obtain an object that is ranked higher than the one you would have ended up with if you had just ranked the objects in the order you prefer them. This is because your ranking list does not affect what objects will be available when it is your turn in the picking order to receive an object, and when your turn arrives, the computer will give you the object you have ranked highest among those that are still available.

Example:

Let's go through an example. For our example we will use four objects that are not used in the experiment and a group of four people to make it simpler to understand and ensure we are making no suggestions about how you should rank the objects in the experiment. The example is purely to help you understand the procedure the computer will use.

We will use the names Ann, Bob, Carol, and Dave. The items are a Pizza, Chips, Soda, and Pretzels. Suppose our group members submit the following rankings:

	Ann	Bob	Carol	Dave
1st choice	Chips	Pizza	Pizza	Pretzels
2nd choice	Pizza	Chips	Pretzels	Soda
3rd choice	Soda	Soda	Chips	Chips
4th choice	Pretzels	Pretzels	Soda	Pizza

After submitting their rankings, the computer randomly determines a picking ordering of the group members. Say that the computer randomly orders Ann, Bob, Carol, and Dave as follows:

1st	Bob
2nd	Carol
3rd	Dave
4th	Ann

The computer then determines the allocations using this ordering and the preferences of each group members using the following procedure

1. Bob was ordered first. His top-ranked object is the Pizza. Thus, he is given the Pizza. This indicated in **bold** in the table.

	Bob
1st choice	**Pizza**
2nd choice	Chips
3rd choice	Soda
4th choice	Pretzels

2. Carol was ordered next. The Pizza is gone, so the objects that remain are the Soda, the Chips, and the Pretzels. According to Carol's list, her top choice of these is the Pretzels, so Carol is given the Pretzels.

	Carol
1st choice	~~Pizza~~
2nd choice	**Pretzels**
3rd choice	Chips
4th choice	Soda

3. Dave was ordered third. The Pizza and the Pretzels are gone, so the objects that remain are the Soda and the Chips. According to Dave's list, his top choice between the Soda and the Chips is the Soda, so he is given the Soda.

	Dave
1st choice	~~Pretzels~~
2nd choice	**Soda**
3rd choice	Chips
4th choice	~~Pizza~~

4. Ann is ordered last. The Pizza, Pretzels, and Soda are gone, so Ann receives the Chips.

	Ann
1st choice	**Chips**
2nd choice	~~Pizza~~
3rd choice	~~Soda~~
4th choice	~~Pretzels~~

Procedures:

During the procedure, all you must do is submit **one list ranking all of the objects.** Note that you will not know your place in the picking order when you submit your rankings. The computer will take everyone's list and then determine the picking order. The picking order is entirely random, and is not influenced by the list you submit. As described above, when your turn comes, the computer will consider all of the objects that are still left, and give you the one that you ranked the highest.

After you are allocated your object, **we will ask whether you want to keep the object or exchange it for various amounts of money using the exact same procedures as in Phase I.**

Earnings:

If this phase is selected for payment the computer will randomly select rows just as in Phase I to determine if you will keep the object allocated to you by the procedure or the amount of money.

Practice:

We will now hand out a worksheet with another example of the procedure. Please work through the worksheet and raise your hand when you are finished. We will come over and check your work and help you if there are any mistakes.

Then, there will be an 8 minute practice period on the computer. The practice period will allow you to practice allocating 5 objects to 5 people. You can submit a ranking list and the computer has robots who submit the other 4 ranking lists. While you will not know the random ordering in the actual procedure when you submit your rank list, this practice will allow you to simulate what you would have gotten if you had submitted different lists. You can experiment with different rank lists as many times as you like for the randomly selected picking order. You can also generate a new random picking order, and experiment further.

For the practice round, we recommend that you first think about the order in which you would actually prefer the objects, and take note of the outcome you get for each possible list you try.

We are doing this practice because you will only submit your ranking list once for the actual experiment and we want to make sure everyone fully understands the procedure before we continue to the actual experiment. You do not earn anything for the practice.

Phase II: Boston

Phase II

For this phase, you will be divided into groups of 5 people, and for each group there will be 5 objects. Each member of the group will end up with exactly one of the objects. The object you get will be determined by your preferences for the object, along with the preference of your group members. Each group member will provide a **preference ranking** of the objects. Because there is only one of each object available, if multiple people list a particular object first, we will use a random tie-breaking procedure to determine who gets what. The details of how this is done are described below.

The Task

You will first be shown the 5 objects. Your task for this phase is to **submit a list ranking the objects from most-preferred to least-preferred.**

After collecting everyone's rankings, the computer will randomly assign each group member a number from 1 to 5 that will be used to break ties. Every group member will be assigned a different number. The computer will then allocate an object to each group member in rounds as follows.

- In the **first round**, the computer looks at everyone's **first choices.**
  - If only one person has an object as their first choice, that person is allocated the object.
  - If more than one person lists an object as their first choice, then the computer will give the object to the person who was assigned the <u>lowest number</u> 1-5.
- In the **second round**, only people who did not receive anything in the first round participate. In the second round, the computer looks at everyone's **second choices.**
  - If only one person has an object as their second choice, then that person is allocated the object.
  - If more than one person has an object as their second choice, then the computer will give the object to the person who was assigned the <u>lowest number</u> 1-5.

  The procedure continues in rounds in the same manner, considering only agents who remain and their third choices in round 3, fourth choices in round 4, etc., until everyone is assigned an object.

It is up to you how to rank the objects. The random numbers 1-5 are not assigned until after everyone submits their rankings. They will only be used if the computer needs to break ties.

Example:

Let's go through an example. For our example we will use four objects that are not used in the experiment to make it simpler to understand and ensure we are making no suggestions about how you should rank the objects in the experiment. The example is purely to help you understand the procedure the computer will use.

We will use the names Ann, Bob, Carol, and Dave. The items are a Pizza, Chips, Soda, and Pretzels. Suppose our group members submit the following rankings:

	Ann	Bob	Carol	Dave
1st choice	Pizza	Pizza	Pizza	Pretzels
2nd choice	Pretzels	Chips	Chips	Soda
3rd choice	Chips	Soda	Pretzels	Chips
4th choice	Soda	Pretzels	Soda	Pizza

After collecting the rankings, the computer randomly assigns each group member a number, in this case 1 to 4 because there are 4 group members. Say that this random assignment resulted in the following:

Name	Random Number
Ann	2
Bob	1
Carol	3
Dave	4

Round 1

- The computer considers the first choice of every group member (indicated in gray in the table below).
- Only Dave has the pretzels as his first choice, so Dave gets the pretzels.
- Ann, Bob, and Carol have the pizza as their first choice.
- Bob has a lower random number (1) than Ann (2) or Carol (3). Bob gets the pizza.
- Ann and Carol get nothing in this round.
- The objects that are assigned are denoted in **bold** in the table below.

	Ann	Bob	Carol	Dave
1st choice	Pizza	**Pizza**	Pizza	**Pretzels**
2nd choice	Pretzels	Chips	Chips	Soda
3rd choice	Chips	Soda	Pretzels	Chips
4th choice	Soda	Pretzels	Soda	Pizza

Round 2

- Only Ann and Carol participate in Round 2. The remaining objects are the Chips and the Soda.
- The computer considers Ann and Carol's second choices.
- Carol's second choice is the chips. Carol receives the chips.
- Ann's second choice is the pretzels, but the pretzels were already taken by Dave in round 1.

- Ann does not receive anything in this round.

	Ann	Carol
1st choice	~~Pizza~~	~~Pizza~~
2nd choice	~~Pretzels~~	**Chips**
3rd choice	Chips	Pretzels
4th choice	Soda	Soda

Round 3

- Only Ann remains. The computer considers Ann's third choice.
- Ann's third choice is the chips, but the chips were already taken by Carol in Round 2.
- Ann receives nothing in this round.

	Ann
1st choice	~~Pizza~~
2nd choice	~~Pretzels~~
3rd choice	~~Chips~~
4th choice	Soda

Round 4

- Only Ann remains. The computer considers Ann's fourth choice.
- Ann's fourth choice is the soda.
- Ann receives the soda.

	Ann
1st choice	~~Pizza~~
2nd choice	~~Pretzels~~
3rd choice	~~Chips~~
4th choice	**Soda**

Procedures:

During the procedure, all you must do is submit **one list ranking all of the objects.** The computer will take everyone's list and determine assignments in each round based on the rank lists following the above procedure, breaking ties using the randomly assigned numbers.

After you are allocated your object, **we will ask whether you want to keep the object or exchange it for various amounts of money using the exact same procedures as in Phase I.**

Earnings:

If this phase is selected for payment the computer will randomly select rows just as in Phase I to determine if you will keep the object allocated to you by the procedure or the amount of money.

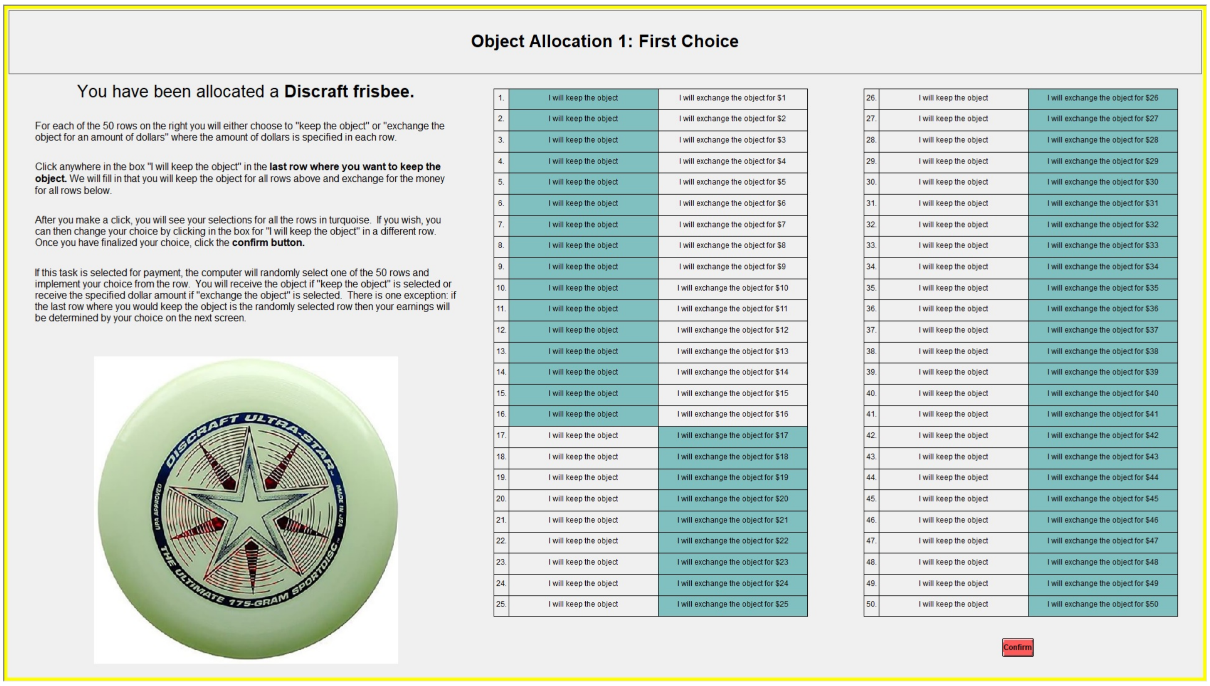
Practice:

We will now hand out a worksheet with another example of the procedure. Please work through the worksheet and raise your hand when you are finished. We will come over and check your work and help you if there are any mistakes.

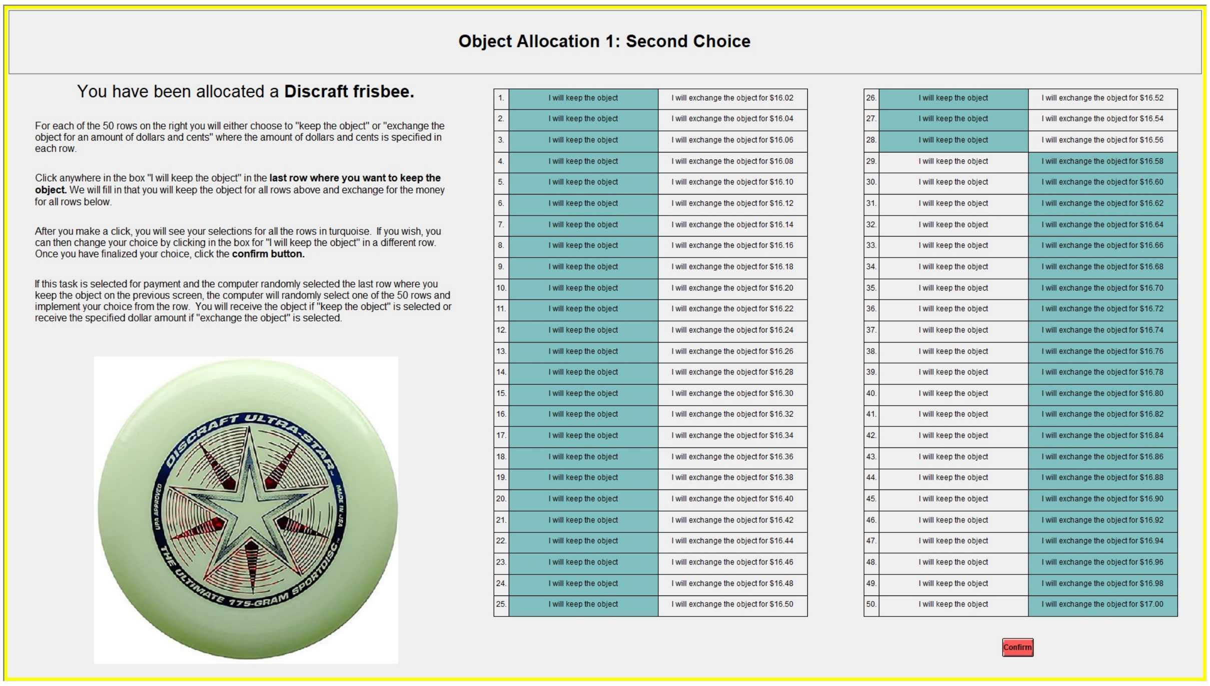
Then, there will be an 8 minute practice period on the computer. The practice period will allow you to practice allocating 5 objects to 5 people. You can submit a ranking list and the computer has robots who submit the other 4 ranking lists. While you will not know the random ordering in the actual procedure when you submit your rank list, this practice will allow you to simulate what you would have gotten if you had submitted different lists. You can experiment with different rank lists as many times as you like for the randomly selected picking order. You can also generate a new random picking order, and experiment further.

For the practice round, we recommend that you first think about the order in which you would actually prefer the objects, and take note of the outcome you get for each possible list you try.

We are doing this practice because you will only submit your ranking list once for the actual experiment and we want to make sure everyone fully understands the procedure before we continue to the actual experiment. You do not earn anything for the practice.



(a) First Screen



(b) Second Screen

Figure 2: Value Elicitation Screen Shots

References

Abdulkadiroğlu, A., Y.-K. Che, and Y. Yasuda (2015): "Expanding "choice" in school choice," *American Economic Journal: Microeconomics*, 7, 1–42.

Abdulkadiroğlu, A., Y.-K. Che, and Y. Yasuda (2011): "Resolving Conflicting Preferences in School Choice: The "Boston" Mechanism Reconsidered," *American Economic Review*, 101, 399–410.

Abdulkadiroğlu, A. and T. Sönmez (2003): "School Choice: A Mechanism Design Approach," *American Economic Review*, 93, 729–747.

Ashlagi, I. and Y. A. Gonczarowski (2018): "Stable matching mechanisms are not obviously strategy-proof," *Journal of Economic Theory*, 177, 405–425.

Bade, S. and Y. Gonczarowski (2017): "Gibbard-Satterthwaite Success Stories and Obvious Strategyproofness," .

Bo, I. and R. Hakimov (2019): "Pick-an-Object Mechanisms," Presentation at WZB Designing and Evaluating Matching Markets Conference.

Chen, L. and J. S. Pereyra (2019): "Self-selection in school choice," *Games and Economic Behavior*, 117, 59–81.

Chen, Y. and T. Sönmez (2006): "School Choice: An Experimental Study," *Journal of Economic Theory*, 127, 202–231.

Cooper, D. J. and H. Fang (2008): "Understanding overbidding in second price auctions: An experimental study," *The Economic Journal*, 118, 1572–1595.

Dreber, A. and M. Johannesson (2008): "Gender differences in deception," *Economics Letters*, 99, 197–199.

Dreyfuss, B., O. Glicksohn, O. Heffetz, and A. Romm (2022): "Deferred Acceptance with News Utility," .

Dreyfuss, B., O. Heffetz, and M. Rabin (2019): "Expectations-based loss aversion may help explain seemingly dominated choices in strategy-proof mechanisms," Tech. rep., National Bureau of Economic Research.

Dur, U., R. G. Hammond, and T. Morrill (2018): "Identifying the harm of manipulable school-choice mechanisms," *American Economic Journal: Economic Policy*, 10, 187–213.

Erat, S. and U. Gneezy (2012): "White lies," *Management Science*, 58, 723–733.

Ericson, K. M. and A. Fuster (2014): "The endowment effect," *Annu. Rev. Econ.*, 6, 555–579.

Featherstone, C. R. (2020): "Rank efficiency: Investigating a widespread ordinal welfare criterion," Working paper.

Featherstone, C. R. and M. Niederle (2016): "Boston versus deferred acceptance in an interim setting: An experimental investigation," *Games and Economic Behavior*, 100, 353–375.

Fischbacher, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental economics*, 10, 171–178.

Fragiadakis, D. E. and P. Troyan (2019): "Designing mechanisms to focalize welfare-improving strategies," *Games and Economic Behavior*, 114, 232–252.

Hakimov, R. and D. Kübler (2021): "Experiments on centralized school choice and college admissions: a survey," *Experimental Economics*, 24, 434–488.

Hassidim, A., A. Romm, and R. I. Shorrer (2021): "The limits of incentives in economic matching procedures," *Management Science*, 67, 951–963.

Holt, C. A. and S. K. Laury (2002): "Risk aversion and incentive effects," *American economic review*, 92, 1644–1655.

Kloosterman, A. and P. Troyan (2020): "School choice with asymmetric information: Priority design and the curse of acceptance," *Theoretical Economics*, 15, 1095–1133.

Köszegi, B. (2006): "Ego utility, overconfidence, and task choice," *Journal of the European Economic Association*, 4, 673–707.

Kőszegi, B. and M. Rabin (2006): "A model of reference-dependent preferences," *The Quarterly Journal of Economics*, 121, 1133–1165.

Li, S. (2017): "Obviously Strategy-Proof Mechanisms," *American Economic Review*, 107, 3257–87.

Meisner, V. (2021): "Report-dependent utility and strategyproofness," *Available at SSRN 3888389*.

Meisner, V. and J. von Wangenheim (2021): "School choice and loss aversion," *Available at SSRN 3985777*.

Ortega, J. and T. Klein (2022): "A More Efficient and Egalitarian Mechanism for School Choice," *arXiv preprint arXiv:2204.07255*.

Pais, J. and A. Pintér (2008): "School Choice and Information: An Experimental Study on Matching Mechanisms," *Games and Economic Behavior*, 64, 303–328.

Pycia, M. and P. Troyan (2020): "A Theory of Simplicity in Games and Mechanism Design," CEPR Discussion Paper 15463.

Rees-Jones, A. and S. Skowronek (2018): "An experimental investigation of preference misrepresentation in the residency match," *Proceedings of the National Academy of Sciences*, 115, 11471–11476.

Shorrer, R. I. and S. Sóvágó (2018): "Obvious mistakes in a strategically simple college admissions environment: Causes and consequences," *Available at SSRN 2993538*.

Troyan, P. (2012): "Comparing school choice mechanisms by interim and ex-ante welfare," *Games and Economic Behavior*, 75, 936–947.

——— (2019): "Obviously Strategy-Proof Implemenation of Top Trading Cycles," *International Economic Review*, 60.

——— (2022): "Non-Obvious Manipulability of the Rank-Minimizing Mechanism," *arXiv preprint arXiv:2206.11359*.

Troyan, P. and T. Morrill (2020): "Obvious manipulations," *Journal of Economic Theory*, 185, 104970.

Zhang, L. and D. Levin (2017): "Partition Obvious Preference and Mechanism Design: Theory and Experiment," .