ABSTRACT

Title of dissertation: ESSAYS IN BEHAVIORAL ECONOMICS

Ian Chadd
Doctor of Philosophy, 2019

Dissertation directed by: Professor Erkut Ozbay
Department of Economics

Chapter 1: In many settings, it is natural to think of limited consideration exhibiting spillovers: attention paid to a particular alternative may "spill over" to another alternative based on shared characteristics, complementarities, features of the choice environment, etc. However, it is not straightforward whether, given choice data, a) preferences among alternatives can be revealed, or b) the network of consideration spillovers can be revealed. Using a novel laboratory experiment, I test a deterministic Network Choice model proposed in previous work and find a plethora of violations thereof, even at the individual level. I then propose a stochastic model, Random Network Choice, and analyze its properties regarding the formation of consideration sets. When applied to the laboratory data, I find considerable consistency with the general Random Network Choice model. Armed with a model of network choice consistent with my experimental data, I consider one application in the realm of advertising to show that such a generalization of so-called "positive spillovers" in attention is necessary to avoid misleading welfare analysis.

Chapter 2: This paper experimentally investigates the effect of introducing un-

available alternatives and irrelevant information regarding the alternatives on the optimality of decisions in choice problems. We find that interaction between the unavailable alternatives and irrelevant information regarding the alternatives generates suboptimal decisions. Irrelevant information in any dimension increases the time costs of decisions. We also identify a pure "preference for simplicity" beyond the desire to make optimal decisions or minimize time spent on a decision problem. Our results imply that the presentation set, distinct from the alternative set, needs to be a part of decision making models.

Chapter 3: To what extent does positive reciprocity extent to environments with uncertainty? In order to answer this question, we propose a new game, the Stochastic Gift Exchange game (SGE), that extends the standard sequential deterministic Gift Exchange game (DGE) into an environment with uncertainty. SGE shares the unique subgame perfect Nash equilibrium with DGE wherein no players trade, leading to a suboptimal ex-ante allocation. However, contrary to DGE, leading models of reciprocity do not predict departures from this equilibrium in the direction of positive giving. When we conduct SGE in a laboratory experiment, we find positive Wages and Effort, indicating the presence of an ex-ante reciprocal motive. Moreover, Wages are lower in SGE than in DGE, indicating both that ex-post reciprocal motives also matter and that laboratory studies of gift exchange, which have been exclusively conducted with DGE, may overestimate the amount of positive reciprocity in the real world. Finally, we conduct two alterations of the SGE to investigate to what extent the source of uncertainty matters for reciprocal giving. Results from these treatments indicate that a) the source of uncertainty does not matter for Wage and Effort determination, but that b) there is evidence of an endowment effect in the ex-ante vs ex-post fairness domains. When endowed with the ability to affect the ex-ante (ex-post) allocation, ex-ante (ex-post) reciprocal motives dominate. Such a phenomenon runs contrary to the additive separability of ex-ante and ex-post motives, a common assumption in leading models that incorporate both risk and social preferences. Our results suggest new directions for future theoretical explorations of ex-ante reciprocity.

Essays in Behavioral Economics

by

Ian Chadd

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Professor Erkut Ozbay, Chair/Advisor
Associate Professor Emel Filiz-Ozbay, Co-Advisor
Professor Yusufcan Masatlioglu, Co-Advisor
Professor Daniel Vincent
Associate Professor Michael Kimbrough (BGMT), Dean's Representative

Dedication

To Patrick, for everything

Acknowledgments

I gratefully acknowledge everyone who made this thesis possible.

First, I owe a debt of gratitude to my advisor Dr. Erkut Ozbay. Without his valuable advice and depth of knowledge, this thesis would not be what it is today. The experience I gained collaborating with him on Chapters 2 and 3 of this thesis was invaluable. To that end I also thank Dr. Emel Filiz-Ozbay. Her attention to detail and rigorous demands for precise analysis made for a productive collaborative process in the development of Chapter 2 and for meaningful advice throughout the development of Chapter 1. I also thank Dr. Yusufcan Masatlioglu for helpful guidance in the development of the theory portion of Chapter 1. The additional members of my committee, Dr. Daniel Vincent and Dr. Michael Kimbrough, must be acknowledged for their willingness to serve and review the manuscript of this thesis.

I'm grateful to my many peers in the Department of Economics. My work no doubt benefitted from numerous conversations with Prateik Dalmia. I'd like to recognize Veronika Penciakova in particular for her ceaseless provision of support over my years at UMD. This thesis would never have been completed were it not for shared writing sessions with her as we held each other accountable for our work.

I acknowledge generous financial support from the National Science Foundation in the form of Doctoral Dissertation Research Improvement Grant. These funds provided for the subject incentive payments for Chapter 1 of this thesis and for travel support in the initial stages of my job market in the fall of 2018.

Finally, I would like to thank the support staff of the Department of Economics, including, but not limited to, Vickie Fletcher, Emily Molleur, Terry Davis, Mark Wilkerson, and Heather Nalley. Their ceaseless efforts to support the faculty and graduate students in the department do not go unnoticed. In particular, I'd like to thank Amanda Statland, with whom I worked closely on the many logistics of running experiments in the Experimental Economics Laboratory. None of the work in this thesis could have been accomplished without her.

Table of Contents

List of Tables

List of Figures

Chapter 1: Random Network Consideration: Theory and Experiment

1.1 Introduction

In decision environments with a large number of alternatives, decision makers (DMs hereinafter) may structure search according to a network of connections between these alternatives. For example, a shopper on Amazon.com utilizes a list of "suggested items" to navigate between available goods. The network of connections between options need not be exogeneously provided by some firm, however. Consider a DM who is considering donating to some charity from the set {Animal Shelter in DC, Animal Shelter in NYC, Homeless Shelter in NYC}. Then a DM who initially considers donating to the Animal Shelter in DC may subsequently consider donating to the Animal Shelter in NYC because of the shared attribute of being an Animal Shelter. Similarly, the same DM may then consider the Homeless Shelter in NYC because of the shared geographic location with the Animal Shelter in NYC. The DM will eventually consider both the Animal Shelter in DC and the Homeless Shelter in NYC, even though the two charities share no common attributes. If attention "spills over" between options in this manner in some decision making environment, it would be important for firms to be able to properly elicit the network from the choices of DMs and attention data, if observable.

Indeed, there is evidence from the marketing body of literature to suggest that DMs exhibit such attention spillovers. Shapiro (2018) shows that Direct-To-Consumer advertising exhibits positive spillovers in the case of pharmaceutical anti-depressants: sales of a given drug increase by about 1.6% in response to the advertisement of a rival drug. Sahni (2016) provides experimental evidence that suggests that these positive spillovers are indeed attention-based by studying the response to online advertising in the restaurant market. Advertising a particular restaurant online can increase sales leads[1] to a competing restaurant by round 4%. Finally, Lewis and Nguyen (2015) show that online advertisements can lead to an increase in online searches for competitors' brands by up to 23%.

These marketing studies on attention spillovers have been focused on brand or product categories: the advertisement of a particular good has an effect on consideration of all goods in the same category as that which is advertised. However, two stylized facts suggest that this implicit modelling restriction may be too strong: i) Shapiro (2018) finds a variety of advertising elasticities between goods even in the same defined "category," and ii) Sahni (2016) finds differential effects of rival advertising based on features of the firm (e.g. firm age, aggregate review scores, etc.). A general model of attention spillovers would then need to represent such spillovers as operating on a network of connections between options, with this category-specific

[1] Sales leads are defined in Sahni (2016) as the consumer searching for the restaurant's phone number, which is observable in his dataset.

treatment as a special case.

Beyond the importance that a model of such network consideration has for firms, a precise model of network consideration is important for welfare analysis. Indeed, a common refrain among these marketing studies of attention spillovers is that these positive externalities lead to an under-allocation of advertising relative to the social optimum; Shapiro (2018) presents a supply-side model to make this case. However, if consideration is modelled as following a more general network structure, this is not necessarily true. I show this much in the Section 3.4.

In this work, I present the results of an experiment designed to test the consideration set properties of several nested models of network consideration. To my knowledge, this is the first experimental study of a decision (i.e. non-strategic) environment with a network structure. First, the deterministic special case, studied previously by Masatlioglu and Suleymanov (2017) and which I'll call "Network Choice" (NC hereinafter), is leveraged to structure the parameterization of the laboratory experiment. NC also serves as a deterministic baseline model against which to test the elicited attention data. The consideration set properties of NC are quite strong and I find evidence that attention, even at the subject level, is not consistent with NC in the observed data. In light of the pervasiveness of violations of NC, I suggest a more general stochastic model, which I'll call Random Network Choice (RNC hereafter). This model shares several features with NC. First, it exhibits limited consideration whereby the DM only considers a subset of the available set of alternatives. It also possesses a form of status quo bias where the status quo or "starting point" of the DM determines the set of alternatives that are reachable ac-

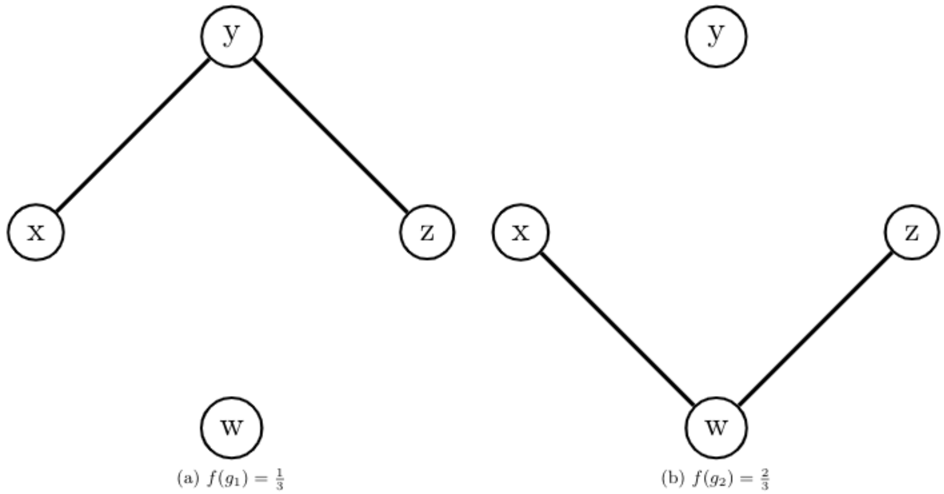


Figure 1.1: Example random graph on four options

cording to the random network structure. However, RNC utilizes a general random product network structure, as opposed to a deterministic network, as is assumed in NC. The generalization to a random network structure allows for more general consideration set mappings and better fits the experimental data. To preview how this model works, consider the following example:

In the above example, the options $\{w, x, y, z\}$ are connected to one another in a random graph structure. The random graph is represented as a distribution over the set of all possible graphs on these four options. In Figure 1.1, the graph g_1 occurs with probability $\frac{1}{3}$, and the graph g_2 occurs with probability $\frac{2}{3}$. All other possible graphs occur with probability 0. Consider a DM who starts at option x and considers options according to RNC. Then with probability $\frac{1}{3}$, the network g_1 is in effect, in which case attention spills over from option x to option y, then from option y to option z. Let $\Gamma_x(T \mid S)$ be the probability that set T is considered when S is available and the starting point is x. Since there are no other networks that connect the set $\{x, y, z\}$, $\Gamma_x(\{x, y, z\} \mid \{w, x, y, z\}) = f(g_1) = \frac{1}{3}$. Notice that, similarly, $\Gamma_x(\{w, x, z\} \mid \{w, x, y, z\}) = f(g_2) = \frac{2}{3}$. The RNC model works as follows: given a starting point x and an available set S, the DM forms stochastic consideration sets according to some distribution over possible networks. From each consideration set, the DM chooses the option that is maximal according to some partial order $\succ$. The NC deterministic model is a special case of RNC where $f(g) = 1$ for some network g, and $f(g') = 0$ for all other networks g'.

Given some dataset that is consistent with RNC, the properties of RNC are such that it may admit an infinite number of representations. In some settings, this may not be a desirable property. For this reason, I then consider a special case of RNC, which I dub the "Pseudo-Markovian RNC" model (PM-RNC hereinafter) due to its proximity to "Markov networks," common in the network analysis body of literature.[2] Under PM-RNC, I show that the PM-RNC representation of some set of stochastic choice data must be unique (up to permutations of preferences between alternatives for which preferences cannot be revealed with the given dataset). I present an additional necessary consideration set mapping property for PM-RNC, Binary Separability, and take this to the experimental data. I find mixed evidence of consistency with Binary Separability, suggesting that there is likely a family of RNC special cases between the most general RNC model and PM-RNC that i) adds structure beyond RNC in the direction of PM-RNC, but ii) is similarly consistent with my experimental data. An exploration of such classes of models is beyond the scope of the current work and would make for a fruitful next step in the study of

[2] See Frank and Strauss (1986) for a discussion of Markov networks.

network consideration.

This paper proceeds as follows. Related literature, both theoretical and experimental studies, are reviewed in Section 1.2. The experimental design and results of tests of NC are presented in Sections 1.3 and 1.4. Section 1.5 presents the RNC and PM-RNC models which are tested in Section 1.6. These results are discussed in light of an application to advertising in Section 3.4. Section 3.5 concludes.

1.2 Related Literature

1.2.1 Experiments

The experiment contained herein is most closely connected to a growing body of literature in economics on experimental investigations of limited attention. Firstly, this experiment elicits data regarding consideration sets in a manner complementary to earlier work. Reutskaja et al. (2011) rely on eye-tracking technology to infer the consideration and search behavior of subjects.[3] Caplin et al. (2011) elicit choice process data as defined previously in Caplin and Dean (2011). Instead of directly observing consideration through eye-tracking technology, Caplin et al. (2011) incentivize the revelation of the path of present-best options at each point in time during which the subject is evaluating a set of options. Geng (2016) studies the impact of a status quo on attention allocation as measured by decision and consideration time. Finally, Gabaix et al. (2006) use the MouseLab coding language to investigate subject attention in a setting with attribute-level information regarding

[3]See Orquin and Loose (2013) for a review of eye-tracking studies in decision making.

available options.

Several studies of attention and information acquisition have been devoted to testing, estimating, or informing theoretical models. Dean and Neligh (2017a) present a set of experiments regarding the rational inattention model of Sims (2003; 2006), generalized in Caplin and Dean (2015), where they document consistency with a generalized model beyond the Shannon case. Chadd et al. (2018) show that the presentation of irrelevant information can affect the consideration set in a manner not predicted by extant models of limited attention. The aim of this experiment is similar to these previous studies in that it seeks to determine consistency with a model of network consideration formation.

The experimental body of literature on networks is often focused on environments where the nodes on the network are optimizing agents and not feasible options to be considered by a central DM. A number of studies exist of network games, where agents are connected to one another via a network structure (See Charness et al. (2014) for a canonical example and Choi et al. (2015) for a thorough survey of such experiments through 2015). In a similar vein, more recent studies have been focused on dynamic network formation, in which agents enter a network sequentially and choose to connect themselves to a subset of extant nodes (agents) in the network. Neligh (2017) shows that entrants to a network "vie for dominance" by connecting to many extant nodes in a manner consistent with forward-looking behavior.

While the experiments above on network formation and network games are at least nominally related to the experiment contained herein insofar as they are explorations of "networks" in economic settings, their connection to the current experiment ends there. All of the above are game-theoretic explorations of behavior in network structures, whether they be exogenously determined or endogenously determined in equilibrium. RNC and PM-RNC are both decision theoretic models and involve no strategic interaction between multiple agents.

1.2.2 Theory

The proposed RNC model contained herein is closely related to several models of path-dependent attention and choice. Masatlioglu and Suleymanov (2017) present a model of Network Choice (NC hereinafter) where attention spills over between options in a given deterministic network. In the realm of stochastic path-dependent models of limited consideration, Suleymanov (2018) presents a Path-Dependent Consideration model that is also similar to RNC, in that consideration follows a path of connections between available options according to some stochastic process. Suleymanov (2018) assigns probabilities in this model to <u>paths</u> with some initial starting point, where RNC assigns probabilities to more general <u>networks</u>. Further, Suleymanov (2018) builds on earlier work contained in Masatlioglu and Nakajima (2013) where new elements are added to the consideration set only when they dominate everything that has already been included in the consideration set. RNC does not share this feature, instead allowing consideration sets to evolve stochastically, independent of the preference relation. The same approach is used in NC, though for a deterministic setting.

Several other models of limited consideration are based on stochastically determined consideration set mappings. Manzini and Mariotti (2014) first explore consideration sets that are stochastically determined. In contrast to RNC, the model of Manzini and Mariotti (2014) focuses on consideration of individual options in the feasible set where each feasible option is considered with some fixed probability. This results in choice probabilities that violate a regularity condition of Luce (1959), where adding an element to the feasible set should not increase the choice frequency of a given element previously available. In more recent work, Cattaneo et al. (2017) present a "Random Attention Model" (RAM hereinafter), that actually relies on violations of the Luce regularity condition to reveal preference. They apply a monotonicity condition on attention rules of the following form:

$$\text{For any } a \in S - T,\ \Gamma(T \mid S) \leq \Gamma(T \mid S - a).$$

where $\Gamma(T \mid S)$ is the probability that the set T is considered when S is available. In Section 1.5, I show that RNC satisfies a starting-point contingent version of this monotonicity condition. This allows me to directly connect the revealed preference approach in Cattaneo et al. (2017) to that in RNC.[4]

RNC also shares features with a number of models that exhibit status quo bias. Note that, in accordance with the distribution over networks of options, a change in the starting point may change both consideration probabilities and, subsequently, choice in RNC. In this way, RNC exhibits a form of status quo bias akin to that explored in Masatlioglu et al. (2005), Masatlioglu and Ok (2013), and Dean

[4]Manzini and Mariotti (2014), Suleymanov (2018), and Cattaneo et al. (2017) are not the only examples of random attention models. See Cattaneo et al. (2017) for a full review of random attention models and their connection to RAM, of which RNC is a starting-point contingent special case.

et al. (2017). However, in the models presented in Masatlioglu et al. (2005) and Masatlioglu and Ok (2013), the status quo affects what is considered by the DM according to whether the status quo dominates an option, with only undominated options being considered. The status quo rules out consideration of certain options more generally in Dean et al. (2017). In contrast, the status quo (or starting point) of RNC simply affects which networks of connections may feasibly be followed in the DM's search - an assumption that is independent of preferences and which is undefined in the absence of a status quo (starting point).

1.3 Experiment

In order to test the deterministic NC model, we construct a laboratory environment with several goals. First, the environment must mimic a choice setting where distinct options are linked to one another via a product network. Second, the environment must induce the subject to behave as if they were in the real world analogue to the laboratory environment - that is, choice must be properly incentivized. Finally, we err on the side of creating an overly restrictive environment in order to test the NC model where it is most likely to succeed. That is, if NC fails in this context, it is not likely to succeed in a real world analogue with more complicated considerations or fewer restrictions.

1.3.1 General Environment

A total of 107 undergraduate subjects at the Experimental Economics Laboratory at University of Maryland, College Park participated in this experiment across eight sessions. On average, subjects earned \$23.63 for approximately 90 minutes of time spent in the lab.[5]

It is helpful to consider the experimental environment from the perspective of a given subject. The subject faces 31 distinctive extended problems, each defined by a starting point x and a set of available options, S, just as in the theory. For each extended decision problem (x, S) the subject's task is to select the option with the highest value among the ones they consider. For each extended decision problem, the subject's payoff is simply the value of the option they have chosen, converted to cash. While subjects make decisions in each of 31 extended decision problems, they are only paid for one, which is chosen randomly at the end of experiment. Subjects do not know which extended decision problem will be chosen when making decisions, so they are incentivized to treat each decision as if it is the one for which they are paid.

Each option is described by four separate attributes: Shape, Pattern, Size, and Number. The value of an option is simply the sum of the value of its attributes, denominated in Experimental Currency Units (ECUs). Each attributes can take on

[5] A single pilot session was conducted with 16 subjects in order to test the experimental software and receive feedback regarding clarity of the instructions used. A few minor changes were made to the design and instructions following this pilot experiment, including, but not limited to, the use of extended decision problem unique option labels. These subjects are not included in any of the analysis contained herein.

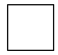
Option X			
Shape	Pattern	Size	Number
		SMALL	5

Table 1.1: Option Example

one of 5 values, from 1 ECU to 5 ECU, resulting in 625 distinct possible options, with values ranging from 5 to 25 ECU. A full table of attribute values can be found in Appendix A.8. For clarity, consider the following option described by these four attributes:

Option X in Table 1.1 is described by 4 attributes: Square, Two-Bar Pattern, Small, and 5. These each pay off 2 ECU, 3 ECU, 2 ECU, and 5 ECU, respectively. Then the value of Option X is 13 ECU ($= 2 + 3 + 2 + 5$). Deciding which option has the highest value in any extended decision problem is thus non-trivial, since it requires i) associating an attribute with its value per the payoff table provided in the instructions and ii) calculating the resultant option value from the sum of its attribute values.

At the start of each extended decision problem, the subject is first shown information for the starting point and no other available option. This information includes an option identification label, unique at the extended decision problem level (i.e. “Option 5” is displayed at the top of the screen when information for Option 5 is presented), attribute information for the displayed option, as well as two lists of information (explained below). In addition, the interface displays information for the subject’s provisional choice at all times (explained in detail in Section 1.3.2).

In order to navigate to information for another option, the subject can utilize two lists on their screen: i) a list of "Linked Options" and ii) a list of "Options Already Viewed." The list of "Options Already Viewed" simply lists the options within the available set for which the subject has already viewed attribute information (defined as having navigated previously to the option information page for that option). To see information for an option other than the one currently displayed, the subject may simply click on the option label in one of these lists and then click a box labelled "View the Selected Option" pertaining to that list. At that time, all relevant information on the screen will update to reflect information for the option to which the subject has navigated.

The list of "Linked Options" displays a list of options that are said to be "linked" to the currently displayed option. An option is said to be "linked" to another if the two share two or more attributes. Thus, for Option X in Table 1.1, if another available option also had the Shape attribute "Square" and the Number attribute "5," it would be included in this list of linked options for Option X. An option that only shared one attribute, but no more, with Option X would not be included in this list. It is through this method that the design induces an exogenous network structure on the set of available options.

This system of "linking" options to one another was chosen for two reasons. First, in order to mimic a real-world environment where NC may be an appropriate model, the experiment necessitated an exogenous network of some form. Second, this particular exogenous network structure was chosen over a more conservative alternative in order to avoid potential subject confusion or experimenter demand effects. In an alternative design where "links" between options are simply agnostic of characteristics of said options, subjects may ask themselves why providing the network structure is necessary in the first place. This may lead to the perception of some deception on the part of the subject or general confusion. The chosen network structure is both easy to understand and mimics real world scenarios where we might believe that NC is the correct model for individual choice.

It is through this navigation process that I argue the current design properly incentivizes revelation of the consideration set for each extended decision problem. In effect, navigating from one option to another "uncovers" hidden information in the extended decision problem regarding the attribute information for each option. Other experiments, both in psychology and experimental economics, use similar designs. I view the design used herein as complementary to approaches incorporating MouseLab, eye-tracking, and choice-process elicitation procedures discussed in Section 1.2.1.

In the baseline version of this experiment, "linked" options were displayed in a list without any additional information regarding these options. For robustness, a variation of this display method was used for half of the sessions. In this variation, the full list of "linked" options was split into four lists, one for each attribute. The option linked to the currently displayed option was then displayed in the lists for the attributes that it shared with the currently displayed option. The goal of using this variation is to determine whether consistency with NC was dependent on arguably minor features of the laboratory environment. In all of the following, whenever statistical tests are conducted separately based on this variation, I use "Baseline" to refer to the original context-less display and "Context" to refer to those observations that came from the variant with more context provided as to the source of the link between options.

Each extended decision problem has a time limit of 75 seconds, and the subject can choose to stop viewing information at any time prior by clicking a “Stop” button located at the bottom of the interface. If this is done, the subject may not view any additional information for options and may not further alter their provisional choice. Stopping the extended decision problem does not allow the subject to immediately move on to the next extended decision problem, however; they must wait for the entirety of the 75 allotted seconds to pass before moving on. This design was chosen to disincentivize haphazard choices on the part of the subject in the interest of finishing the experiment early.

At the end of the experiment, one of the 31 extended decision problems was chosen at random (with each extended decision problem chosen with equal likelihood), and subjects were paid for that single choice only. Once these extended decision problems were completed, they were asked a set of demographic questions on age; gender; self-reported ACT, SAT, and GPA scores; native language; and major of study. They were also given the opportunity to explain their decisions and indicate whether they felt they sufficiently understood the instructions to the experiment.

1.3.2 Choice Process Data

The experimental design elicits choice process data a la Caplin et al. (2011) in the following manner. Choices in each extended decision problem are treated as provisional, in that choosing a new option does not end the current period. This simply updates the subject's provisional choice, allowing the subject to make a number of switches between provisionally chosen options within a single period. At all times, information regarding the subject's provisional choice was displayed in the upper-right portion of the experimental interface, including the option label (e.g "Option 12") for the provisional choice and attribute information. This information was provided as a reference for the subject to avoid concerns of imperfect recall during the option evaluation process.[6]

At the end of each period, a "decision time" was chosen randomly from a uniform distribution from 2 to 75 seconds. The provisional choice held by the subject at the realized decision time was then treated as the final choice for the extended decision problem and subjects were paid the value of the option held at that time.

In each period, while subjects were initially shown the information for the starting point, they did not initially have any option provisionally chosen. They must then choose some option to serve as their initial provisional choice (usually the

[6]It should be noted that, while the appearance of this information differs between the design used herein and that used in Caplin et al. (2011), the two share the feature of always displaying the subject's provisional choice. In Caplin et al. (2011), the provisional choice is indicated by a selected row in a list of continuously displayed available options. In the experiment contained herein, this information is contained in a portion of the interface that simply updates when the provisional choice is altered by the subject.

starting point itself). For this reason, the lower end of the decision time support was 2 seconds, giving the subject time to choose an initial provisional choice and thus minimizing the number of observations for which the subject might be paid nothing for a given extended decision problem.

1.3.3 Data Generation Process

For each extended decision problem, both the set of available options and the starting point were chosen intentionally to create explicit tests of properties of consideration sets in NC. This design was chosen to ensure that there would be a sufficient number of tests of each consideration set property. One alternative design would have randomized the extended decision problems presented to subjects. With four attributes, each taking on one of five different values, the grand set of alternatives is of size 625, with $2^{625} - 1$ unique non-empty subsets. With such a large dataset over which to randomize, it would be highly unlikely that the final dataset would end up with a sufficient number of tests of the properties of NC using a reasonable number of laboratory subjects. Extended decision problems were thus chosen such that the observations gleaned from each would constitute, at minimum, one test of some axiom of NC.

1.3.3.1 NC: Upward Monotonicity

The first property of consideration sets in NC that I utilize to create extended decision problems is Upward Monotonicity. For some extended decision problem

(x, S), let $\Gamma_x(S)$ be the set of all options in S which are considered when x is the starting point. Then the NC property of Upward Monotonicity is as follows:

B1. Upward Monotonicity: $\Gamma_x(T) \subseteq \Gamma_x(S)$ for all $T \subseteq S$

In essence, this property describes the process of aggregation across nested extended decision problems. Under the deterministic NC model, if the DM faces (x, T), they will consider all of those options which are "reachable" from the starting point x and are also in T.[7] Then when the DM is confronted with (x, S) for $T \subseteq S$, it should be clear that all those options which were reachable from x and in T remain reachable under (x, S) (i.e. nothing about the underlying connections between options has been changed). Moreover, since $T \subseteq S$, these options are also still available and therefore should still be considered under (x, S).

In order to test this property in the lab, I define five extended decision problems that are "nested" within one another. Let $\delta_i = (x, A_i)$ be one of these five extended decision problems. Each A_i was then chosen such that $A_1 \subseteq A_2 \subseteq ... \subseteq A_5$. The starting point x was chosen such that $x \in A_1$. A violation of Upward Monotonicity would then look like $\Gamma_x(A_i) \nsubseteq \Gamma_x(A_{i+1})$. From these five extended decision problems, we then have 10 separate tests of Upward Monotonicity per subject, or 1070 in total across 107 subjects.

[7] Masatlioglu and Suleymanov (2017) say that y is "reachable" from x if there exists a path from x to y in the network of connected options. I follow in their footsteps and use the same terminology here.

1.3.3.2 Symmetry

The next property of NC to be tested concerns the "undirectedness" of how products are connected in NC. By the definition of "reachable," it should be clear that if y is reachable from x, x is also reachable from y. This is simply the result of consideration spilling over in either direction of a connection between options, regardless of the origin. This has a clear implication for how consideration sets should compare across the same available set, but given distinct starting points, which is captured in the NC Symmetry property:

B2. Symmetry: If $y \in \Gamma_x(S)$, $\Gamma_x(S) = \Gamma_y(S)$ for all S.

To test the Symmetry Axioms of NC, I repeat the available sets in each of δ_i above, letting γ_i denote one such extended decision problem. These Symmetry extended decision problems use a distinct starting point $y \neq x$, with $y \in A_1$. A test of these Symmetry Axioms would then involve a comparison between consideration sets in δ_i and γ_i. Then, in total, these ten extended decision problems create a maximum of five tests of Symmetry for each subject. However, notice that the Symmetry property only applies if $y \in \Gamma_x(S)$, which may not be born out in the data for a given subject. The actual total number of tests of Symmetry will then be endogenously determined by consideration behavior of subjects.

1.3.3.3 Path Connectedness

The final property of NC to be tested concerns the impact of an option that uniquely provides a connection between two other options. This property, Path Connectedness, essentially states that the revelation that some option y is required to make z reachable from x should also reveal i) that y is reachable from x in the absence of z and ii) that z is reachable from y in the absence of x. Formally, Masatlioglu and Suleymanov (2017) write this property as follows:

B3. Path Connectedness: If $z \in \Gamma_x(S)$ and $z \notin \Gamma_x(S \setminus y)$, then $y \in \Gamma_x(S \setminus z)$ and $z \in \Gamma_y(S \setminus x)$

This property is best understood through a simple example. In Figure 1.2, clearly $z \in \Gamma_x(\{x, y, z\})$, but $z \notin \Gamma_x(\{x, z\})$; the only connection between x and z passes through y. Path Connectedness essentially identifies the fact that this tells us two things. First, y must then be connected to x, independent of z. Similarly, y must then be connected to z, independent of x. So, we can then say i) $y \in \Gamma_x(\{x, y\})$ and ii) $z \in \Gamma_y(\{y, z\})$, as stated in the property above.

This property utilizes four separate extended decision problems: (x, S), $(x, S \setminus y)$, $(x, S \setminus z)$, and $(y, S \setminus x)$. Further, note that the hypothesis of the property, similar to that of the Symmetry property, is going to be endogenously determined by subject consideration data: it may be the case that we present both (x, S) and $(x, S \setminus y)$, to the subject and that their behavior does not satisfy the hypothesis of this statement. In order to increase the probability that there is a sufficient number

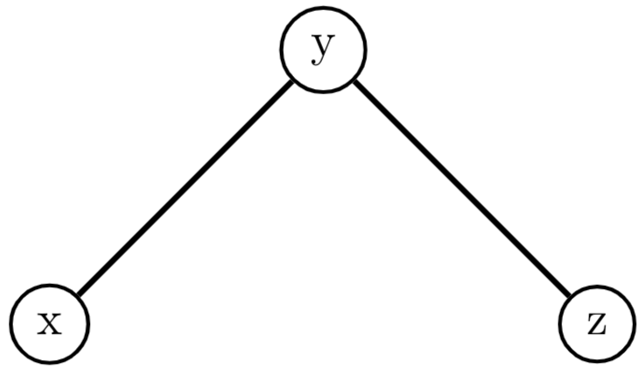


Figure 1.2: Example graph with three options

of observations where the hypothesis is satisfied, two separate options for y and z in the above are presented to each subject, holding x and S fixed. There are thus four possible tests of Path Connectedness for each subject, constructed using seven extended decision problems. The actual number of tests conducted are a function of the consideration set data.

In total, these three properties of NC lead to the creation of 17 extended decision problems to be used in the laboratory experiment. The remaining 14 (out of 31) were constructed to test axioms on choice data of NC, along with the choice axioms of a related model contained in Suleymanov (2018). The results of these tests are not included, as the focus of the main body of the paper is an exploration of consideration set formation in NC.

1.4 Results: NC

The results of tests of consistency with NC are presented below. Given data on both choices and consideration sets, one can check for consistency in two separate ways: simultaneous and sequential. Under a simultaneous test, one would test whether the resultant choices were consistent with some NC representation. This is, in general, the approach taken when consideration sets are unobserved and consistency with some decision theoretic model can only be tested using choice data. In this experiment, consideration sets are elicited, so one can take a sequential approach, paying more attention to the consideration set formation process. In a sequential test consistency with NC is separated into two questions. First, in each extended decision problem, does the subject choose a money-maximizing element of the consideration set? Second, is the formation of random consideration set mappings consistent with an NC representation (i.e. do consideration sets satisfy Upward Monotonicity, Symmetry, and Path Connectedness)? Subsection 1.4.1 presents general experimental results and demographic information. Subsection 1.4.2 answers the first question on optimality of choice. Finally, Subsection 1.4.3 reports tests of the NC properties.

1.4.1 General Results

In all of the below, statistical tests were conducted on aggregate data, pooled across the Baseline and Context displays, except where explicitly mentioned. Tests of differences between the two displays that were omitted in the main text can be found in Appendix A.7. Upon completion of the experimental task, subjects filled out a brief demographic questionnaire which asked questions on Age, self-reported SAT and ACT scores (if any), self-reported GPA, and Gender. Results are presented in Table 1.2.

Table 1.2: Demographic Information

	Age	SAT	ACT	GPA	Female
mean	20.68224	1810.833	30.11429	3.363364	.4485981
sd	1.551616	324.1159	3.946491	.438533	.4996913
min	18	1100	20	2	0
max	27	2360	36	4	1
count	107	84	35	107	107

In order for data on self-reported GPA, SAT, and ACT to be used in subsequent analysis, responses were normalized as in Cohen et al. (1999), Filiz-Ozbay et al. (2016), and Chadd et al. (2018): Let j be the variable under consideration with $j \in \{\text{GPA, SAT, ACT}\}$, μ_i^j be the value of variable j for subject i, μ_{max}^j be the maximum value of j in the subject population, and μ_{min}^j be the minimum value of j in the subject population. Then let $\hat{\mu}_i^j$, the normalized value of variable j for subject i, be defined as follows:

$$\hat{\mu}_i^j = \frac{\mu_i^j - \mu_{min}^j}{\mu_{max}^j - \mu_{min}^j}$$

such that $\hat{\mu}_i^j$ can be interpreted as the measure of j for subject i, normalized by the distribution of j in the subject population. Some subjects were missing one or more measures for $j \in \{\text{GPA, SAT, ACT}\}$, since these measures were self-reported (and some subjects could not recall their scores on one or more of these measures). Additionally, some responses were outside the range of feasible scores (for example, an SAT score of 20). All subjects could reliably self-report a feasible GPA from the range of 2 to 4, so $\hat{\mu}_i^{GPA}$ will be used for any subsequent analysis involving cognitive ability. Normalized scores are reported in Table 1.3 along with an additional measure for Cognitive Score. For some subject i Cognitive Score is taken to be $\hat{\mu}_i^{SAT}$ if the subject reported a feasible SAT score and $\hat{\mu}_i^{ACT}$ if the subjects did not report a feasible SAT score and reported a feasible ACT score. Cognitive Score is lower than $\hat{\mu}_i^{GPA}$ and has higher variance (likely due to more imperfect recall of SAT/ACT scores relative to GPA).

Table 1.3: Cognitive Scores

$\hat{\mu}_{SAT}$	0.564
	(0.257)
$\hat{\mu}_{ACT}$	0.632
	(0.247)
$\hat{\mu}_{GPA}$	0.682
	(0.219)
Cognitive Score	0.579
	(0.256)
Observations	107

1.4.2 Choice Optimality

For the purposes of testing "choice optimality," I say that a subject chose a "highest-valued option" in a given extended decision problem when the option that was last chosen by the subject before the end of the period (i.e. before 75 seconds elapsed) led to the highest possible ECU payoff among those options that the subject considered. Note that this, in general, is not equivalent to standard mistake rate analysis conducted in Caplin et al. (2011) and Chadd et al. (2018), for example. Here, we only say that a "mistake" was made when the subject ended up choosing a lower-valued option than one that was actively considered in the current period.

Note: we view an option as having been "considered" if the subject navigated to its information at some point in the period. Since we are testing for choice optimality in the context of NC, which says that a DM will choose the optimal option in the DM's consideration set, it is natural to define "mistakes" as described above.

Subjects chose the highest-valued option in 85.675% of extended decision problems (Wilcox $p < 0.001$ for $H_0 : \mu = 1$). Given that the overall mistake rate is non-trivial, I further investigate the determinants of choice optimality through several logistic regression specifications. In Table 1.4, the dependent variable is Correct, a binary variable that takes the value 1 if the subject chose the consideration set optimal option in the extended decision problem and 0 otherwise. Context is a binary variable used to indicate whether the observation came from the Context display. Period goes from 1 to 31, indicating the period in which the extended decision problem was completed. CS_N^{ϵ} is the residual generated from an OLS regression of CS_N onto Period and N. Both the size of the available set (N) and the period in which the extended decision problem is conducted affect the size of the consideration set.[8] These residuals are the portion of CS_N left unexplained by N and Period, and they are used in both models to estimate the effect of consideration set size on choice optimality separately from the effects of N, the size of the available set. Female and GPA are defined as above (i.e. GPA is normalized according to the POMP procedure described in Subsection 1.4.1). In both model specifications, marginal effects from a logistic regression are reported, along with robust standard errors clustered

[8] An interested reader can find these results in the Appendix A.7. I replicate a version of the results contained in Reutskaja et al. (2011), that additional available options lead to more options being considered by the DM.

at the subject level.

From Table 1.4, we can see that the prevalence of sub-optimal choice can partly be attributed to learning: each additional period increases the probability that the choice will be consideration set optimal by 0.1 percentage points, resulting in higher rates of sub-optimal choice in earlier periods. Further evidence of this can be seen by looking at the final period only, where 94.33% of observations are consideration set optimal. Additionally, the size of the available set matters; for each additional element added to the set of available options, the probability that the chosen element will be consideration set optimal decreases by 0.953 percentage points. Somewhat surprisingly, the size of the consideration set itself matters - an additional option considered (holding N and Period fixed) decreases the probability of consideration set optimal choice by about 1.3 percentage points. Note that in neither specification do Context, Female, or GPA matter. This brings us to our first two results on consideration set optimality of choice:

Table 1.4: Determinants of Optimal Choice

	Model 1	Model 2
Context	-0.0156	-0.0164
	(0.031)	(0.031)
Period	0.00175**	0.00174**
	(0.001)	(0.001)
CS_N^ϵ	-0.0126***	-0.0129***
	(0.003)	(0.003)
N	-0.00953***	-0.00952***
	(0.001)	(0.001)
Female		-0.0359
		(0.027)
$\hat{\mu}_{GPA}$		0.0358
		(0.080)
Observations	3276	3276

Standard errors in parentheses

Marginal effects from logit regression specifications

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Result 1. *Consideration set choice optimality in the aggregate is broadly consistent with NC:*

- *85.675% of choices are consideration set optimal*

Result 2. *There is a non-trivial number of observations where choices are not consideration set optimal in a manner not predicted by NC:*

- *larger consideration sets decrease the likelihood of consideration set optimality*
- *larger available sets decrease the likelihood of consideration set optimality*

1.4.3 Property Tests

In Subsections 1.4.3.1 - 1.4.3.3, we consider subject-level data to test the deterministic consideration set properties of NC.

1.4.3.1 Upward Monotonicity

The five extended decision problems constructed to test Upward Monotonicity, when repeated twice (in order to test Symmetry - results in Subsection 1.4.3.2) comprise 20 tests of Upward Monotonicity per subject. In the aggregate, 79.8% of these observations were inconsistent with Upward Monotonicity, as seen in Table 1.5. Moreover, an analysis of the CDF of the proportion of Upward Monotonicity violations per subject, displayed in Figure 1.3, reveals that nearly 50% of subjects had more than approximately 80% of their observations in violation of Upward Monotonicity. No subjects had fewer than 20% of their observations in violation of Upward Monotonicity. Taken together, these results suggest that Upward Monotonicity may be too strong an assumption on how consideration sets are formed in the presence of an exogenous network, even at the individual level.

Table 1.5: Aggregate Test of NC Upward Monotonicity

	Monotonicity Violation
Mean	0.798
Std Error	0.000
N	2140

$p < 0.001$ for aggregate test of $H_0 : \mu = 0$

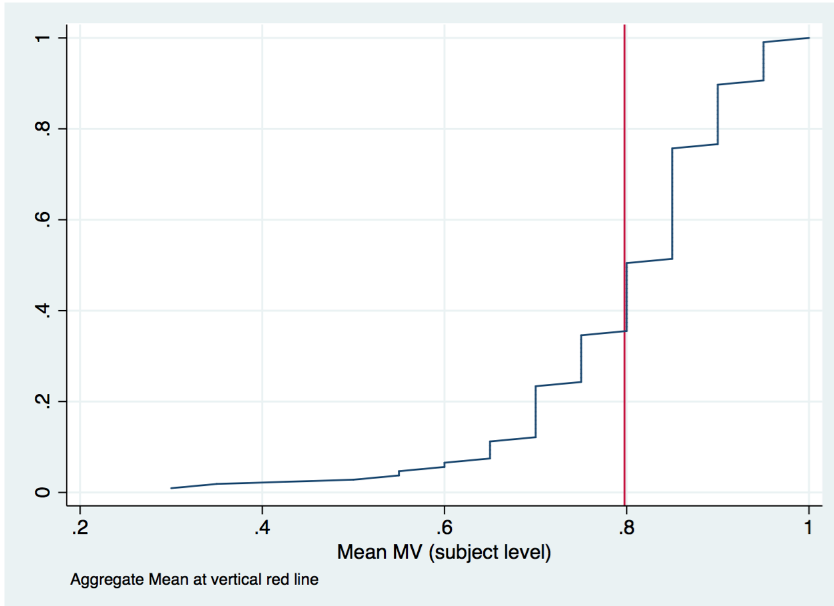


Figure 1.3: Cumulative of Mean Monotonicity Violations by Subject

Given the prevalence of violations of Upward Monotonicity, we ask what characteristics of extended decision problems and the choice environment affect the likelihood of observing a violation. First, note that Upward Monotonicity does not take into account the "distance" between the sizes of the relevant available sets. At first glace, this appears as if it should not matter. If set $\Gamma_x(T)$ is considered when T is available, then this indicates that option x is connected to elements in T through some combination of paths. When $S \supseteq T$ is available, these same paths are still present, so at least all of the elements in T should again be considered. However, if there is some probability that the DM switches from one path or sub-network to another when encountering a new extended decision problem with the same starting point, then the size of $S \setminus T$ may matter.[9] As $S \setminus T$ gets larger, the number of sub-

[9]Note that, in NC, the probability that the DM "switches" consideration sub-networks across identical or nested extended decision problems is implicitly 0. We mention this switching behavior as an empirical possibility in the environments NC is modelling, not as behavior that is consistent with the MS model itself.

networks on S relative to T also gets considerably larger, leading to an increased likelihood that the DM follows some sub-network that is distinct from the one they follow when T is available. We may thus expect that the likelihood of observing a violation of Upward Montonicity to be increasing in $|S \setminus T|$. We see that this is the case in Figure 1.4 below. For extended decision problems (x, A_i), for $A_i \subseteq A_j$, we define the Distance between A_i and A_j to be equal to $|A_j| - |A_i|$. In Figure 1.4, we see that as Distance increases, so too does the likelihood of observing a violation of Upward Monotonicity (though there is considerable overlap in 95% Confidence Intervals for these categories).

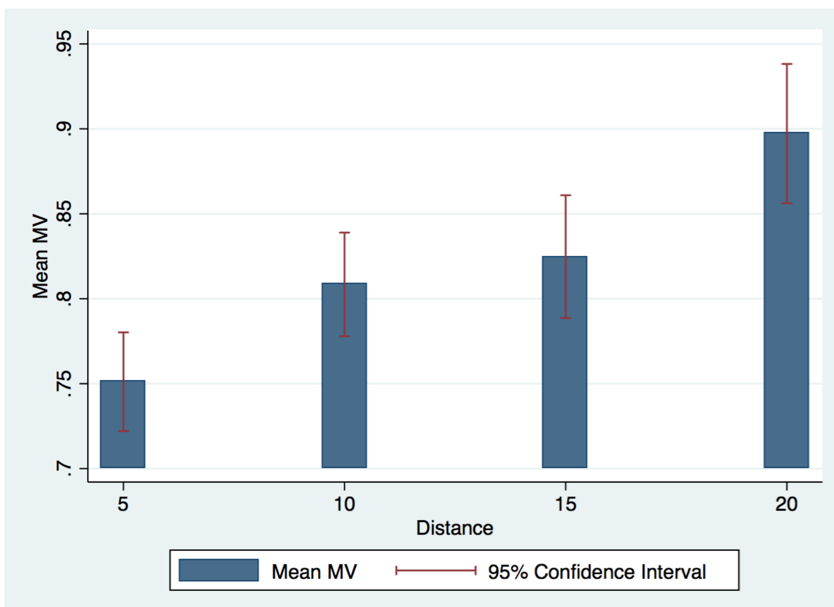


Figure 1.4: MV by $|A_j| - |A_i|$, $A_i \subseteq A_j$

To further investigate the determinants of Upward Monotonicity violations in our data, Table 1.6 reports the results of several logistic regressions. In each model, Context is a dummy variable used to indicate whether the observation came from the Context display, and Female, GPA, and Age are as they were defined previously.

Reported coefficients are marginal effects from logistic regressions and standard errors are robust and clustered at the subject level. From Table 1.6, it initially appears that Distance increases the likelihood of Upward Monotonicity violations by approximately 0.93 percentage points, per the reported marginal effect in Model 1. However, the entirety of this effect in the aggregate is driven by the tests of Upward Monotonicity involving $A_1 \subseteq A_2$. These available sets are of size 5 and 10, respectively. Thus, it appears as if Upward Monotonicity violations are ubiquitous regardless of Distance, provided that the available sets involved are sufficiently large.

Table 1.6: Determinants of Monotonicity Violations

	Model 1	Model 2	Model 3	Model 4
Distance	0.00927***	0.00241	0.00241	0.00241
	(0.00153)	(0.00178)	(0.00178)	(0.00178)
A_1 to A_2		-0.234***	-0.234***	-0.234***
		(0.0259)	(0.0257)	(0.0263)
Context			0.0138	0.0181
			(0.0232)	(0.0223)
Female				0.00377
				(0.0242)
GPA				0.0679*
				(0.0354)
Age				-0.00599
				(0.00586)
Observations	2140	2140	2140	2140

Standard errors in parentheses

Marginal effects from logistic regressions reported

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1.4.3.2 Symmetry

Recall that extended decision problems were constructed such that the five nested available sets, $A_1 \subseteq A_2 ... \subseteq A_5$, were each used in two extended decision problems: (x, A_i) and (y, A_i), $x, y \in A_1$. This results in ten possible tests of symmetry for each subject: for each $(x, A_i) - (y, A_i)$ pair of extended decision problems, we can write two statements of Symmetry to be tested in the data:

$$y \in \Gamma_x(A_i) \implies \Gamma_x(A_i) = \Gamma_y(A_i) \tag{1.1}$$

$$x \in \Gamma_y(A_i) \implies \Gamma_y(A_i) = \Gamma_x(A_i) \tag{1.2}$$

These two conditions are clearly interrelated. If both the hypothesis and implication of condition 1.1 are satisfied for some observation, then so will those of 1.2, and vice versa. However, if the hypothesis is not satisfied in one, it is possible that the other test may fail. In order to rule out double-counting successes (and failures), in all of the following we exclude tests of condition 1.2 unless condition 1.1 is satisfied only trivially (i.e. $y \notin \Gamma_x(A_i)$). We will thus only include a maximum of five tests of symmetry per subject.

Out of a possible maximum of 535 tests of deterministic symmetry, there were 401 observations where the hypothesis of this axiom was satisfied. In the aggregate, 84.29% of these observations violated symmetry (Wilcox sign rank $p < 0.001$ for $H_0 : \mu = 0$). Table 1.7 presents aggregate summary statistics. Hypothesis is a dummy variable indicating whether the observation satisfied at least one hypothesis contained in conditions 1.1 and 1.2, Symmetric is a dummy variable indicating whether the implication in the relevant condition is satisfied (conditional on the hypothesis being satisfied), and Violation is simply equal to 1 - Symmetric.

Table 1.7: NC Symmetry Summary Statistics

	hypothesis	symmetric	violation
Mean	.7495327	.1571072	.8428928
SD	.4336877	.3643564	.3643564
N	535	401	401

At the subject level, we also examine Symmetry violation counts and rates. Of a total of 5 possible tests of Symmetry per subject, subjects satisfied a hypothesis of conditions 1.1 or 1.2 for 3.75, on average. Notably, the maximum number of symmetric observations for a given subject is 3 (out of 5 tests). We can further examine the distribution of Symmetry violation rates in Figure 1.5. Notably, more than half of subjects violated MS Symmetry in 100% of their valid tests.

Table 1.8: NC Symmetry Subject Level Summary Statistics

	Mean	SD	Min	Max	N
Hypothesis N	3.748	(1.237)	0.000	5.000	107.000
Symmetric N	0.589	(0.672)	0.000	3.000	107.000
Violation N	3.159	(1.175)	0.000	5.000	107.000

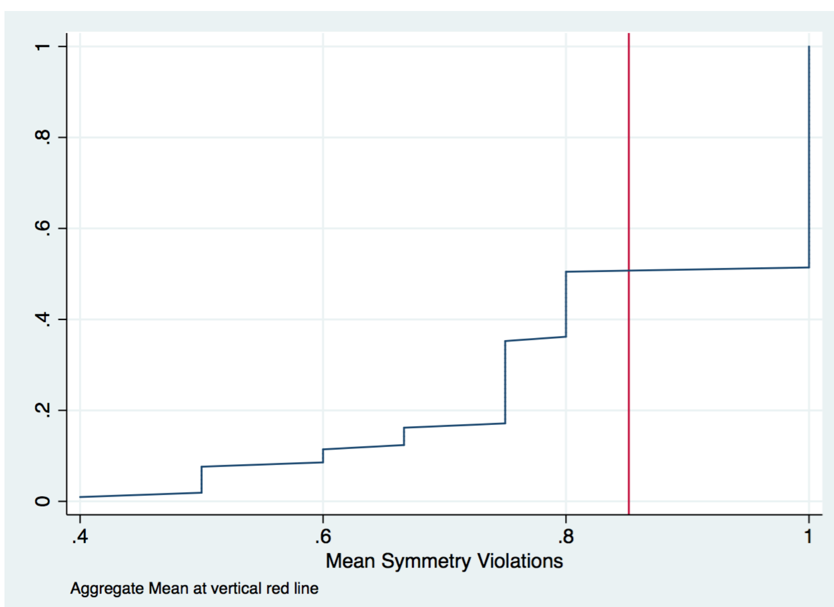


Figure 1.5: Cumulative Mean Symmetry Violations per Subject

We conjecture that one reason Symmetry may fail in this context is that little information is available regarding an individual option prior to its consideration. The deterministic MS Symmetry axiom requires much of the DM: conditional on arriving at some node in the network, the DM follows the same pattern of search for a given available set. If a given subject then follows different paths of consideration starting at, say, option y in (x, A_i) and (y, A_i), then their consideration set will not satisfy symmetry. In the Baseline environment, no information is available to the subject regarding an individual option prior to its consideration. Thus, if we make some information available to the subject prior to an object being considered, this may increase the likelihood of symmetry consideration paths across extended decision problems. We test this hypothesis by comparing the rate of violations of Symmetry in Table 1.9. In the Context environment, 81.9% of observations violated symmetry compared to 87.4% in the Baseline, which runs counter to this hypothesis. However, this difference is not statistically significant (Wilcox $p > 0.10$).

Table 1.9: Symmetry Violations by Context

	Baseline	Context
Mean	0.819	0.874
Std Error	0.026	0.025
N	227	174

Wilcox $p > 0.10$ for $H_0 : \mu_{Baseline} = \mu_{Context}$

While a simple analysis of aggregate Symmetry violations is helpful, it is illuminating to consider subject-level mean violations of symmetry. Recall that we may be including a maximum of five tests of Symmetry for a given subject. Table 1.8 presents summary statistics for subject-level data on the number of tests per subject (Hypothesis) and violations of MS Symmetry at the subject level. Further, from Figure 1.5 we can see that nearly 50% of all subjects violated Symmetry in each observation where the hypothesis was satisfied. When we decompose this cumulative distribution function by informational environment, we can see that a larger proportion of subjects have a 100% Symmetry violation rate in the Context environment than in the Baseline (42.11% vs 58.33%; Mann-Whitney $p < 0.10$).

Therefore, on the whole, subjects tend to exhibit behavior more consistent with Symmetry in the Baseline than when Context is provided, though behavior in both is largely inconsistent with Symmetry.

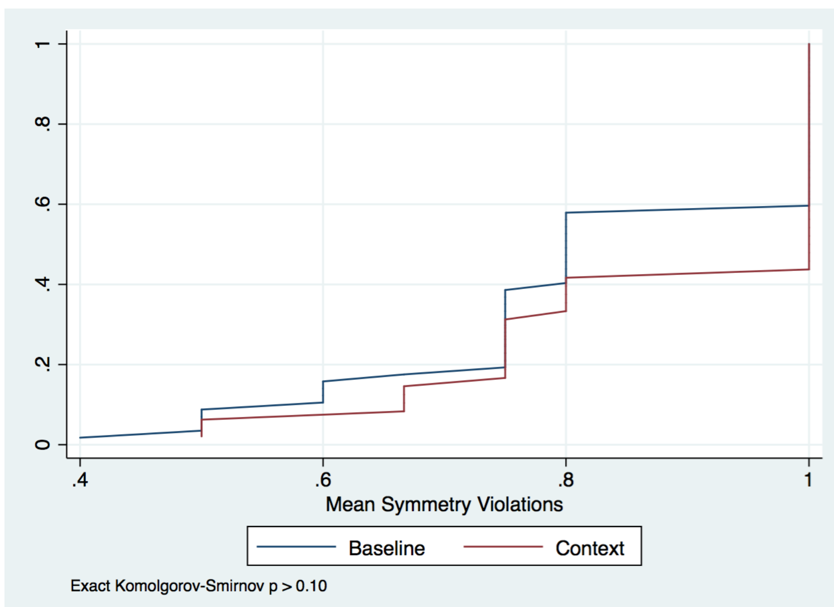


Figure 1.6: Cumulative of Per Subject Mean Symmetry Violations by Context

1.4.3.3 Path Connectedness

With 107 subjects and four possible tests of Path Connectedness per subject, we have a maximum of 428 tests. In the aggregate, only about 23% of possible tests were such that the hypothesis of the MS Path Connectedness axiom was satisfied, making for 99 total tests used, across 67 subjects. Of these 99 tests, roughly 45.5% were consistent with with Path Connectedness, as reported in Table 1.10.

Table 1.10: Aggregate Test of NC Path Connectedness

	Path Connectedness
Mean	0.455
Std Error	0.050
N	99

Wilcoxon signed-rank $p < 0.001$ for aggregate test of $H_0 : \mu = 1$

Thus, taken together, the experimental results pertaining to consideration set data largely reject the NC deterministic model at the subject level. This leads me to the next result:

Result 3. *Consideration set probabilities are largely inconsistent with the deterministic NC model in the experimental data:*

- *Nearly 80% of observations violate Upward Monotonicity*
- *Approximately 84% of observations violate Symmetry*
- *45.5% of observations violate Path Connectedness*

1.5 Random Network Choice

Given that the experimental data is largely inconsistent with the deterministic NC model, I propose a stochastic generalization to be tested against the same dataset. First, I propose the most general Random Network Choice (RNC) model and discuss necessary properties that this model imposes on stochastic consideration set mappings. These properties are directly related to the deterministic properties of NC. The RNC model has a feature that leads to an infinite number of representations for a given set of choice data that is consistent with this model. Guided by the notion that eliciting a unique network structure from choice or consideration set data may be desirable in many applications, I present a special case, Pseudo-Markovian Random Network Choice (PM-RNC hereinafter) which does have a unique representation.

I should note that the following section is meant only to provide suitable modeling alternatives to the NC model that may be consistent with the experimental dataset contained herein. While the NC model of Masatlioglu and Suleymanov (2017) is an axiomatic characterization of choice in the deterministic setting, such an axiomatic characterization of RNC and PM-RNC is beyond the scope of the current work. Instead, I focus on necessary conditions of consideration set mappings to be tested against the experimental data.

1.5.1 Random Network Consideration

Let X be a finite set of alternatives with $\Omega = 2^X \setminus \emptyset$ as the set of all non-empty subsets of X. I consider random networks on X. To that end, let $g = [X, E, \psi]$ be a network consisting of a set of nodes (alternatives) X, edges E, and an incidence function ψ such that $\psi(i, j) \in \{0, 1\}$ indicates whether nodes i and j have an edge between them ($\psi(i, j) = 1$) or not ($\psi(i, j) = 0$). I restrict ψ to be such that $\psi(i, i) = 1$ for all $i \in X$ and $\psi(i, j) = \psi(j, i)$. In other words, I restrict attention in this model to undirected networks on X. In slight abuse of notation, I'll use g_{ij} to refer to the value of $\psi(i,j)$ associated with network g, such that $g_{ij} = 1$ indicates that nodes i and j have an edge between them under g (and vice versa for $g_{ij} = 0$). Let $\mathcal{G}$ be the set of all possible networks on X. Finally, let F be a distribution on $\mathcal{G}$ where I denote the probability that network g occurs by $f(g)$.

To consider subsets $S \in \Omega$, I must restrict attention to only those nodes that are in S. Abusing notation slightly, let $g[S]$ be a node-induced sub-network of g such that $g[S] = [X, E[S], \psi[S]]$ where $E[S]$ is simply the edge-set E minus all edges that involve any nodes $i \notin S$ and $\psi[S]$ is defined as follows:

$$\psi[S](i,j) = \begin{cases} \psi(i,j) & \{i,j\} \subseteq S \\ \\ 0 & \text{otherwise} \end{cases} \tag{1.3}$$

Given a network g, i and j are **connected** under g if there exists an $i - j$ path in g. That is, they are connected if there exists a sequence $(x_0, x_1, ...x_n)$ with $x_0 = i$ and $x_n = j$ where $g_{x_k, x_{k+1}} = 1$ for all x_k and x_{k+1}. Using this terminology, the definition of what it means for a given subset of nodes to be connected under some network g directly follows:

Definition 1. *A network $g \in \mathcal{G}$ is said to be **T - Connected** for some set $T \in \Omega$ if*

1. *t and t' are connected under g for all $t, t' \in T$ with $t \neq t'$*

2. *t and t' are not connected under g for each $t \in T$ and $t' \in X \setminus T$*

In other words, a network g is $T-$Connected if all of the elements in T are connected to one another under g and no element of T is connected to a path that leads out of T.

Given an extended problem (x, S) which consists of an alternative set S and starting point $x \in S$, the DM forms a consideration set stochastically. I begin by defining a general starting-point contingent random consideration set mapping that gives the probability that the consideration set is T, given that the starting point is x under alternative set S:

Definition 2. *A function $\Gamma_x : \Omega \times \Omega \to [0, 1]$ is a starting-point contingent random consideration set mapping if the following is true:*

$$\Gamma_x(T \mid S) = \begin{cases} 1 & \text{if } \{x\} = T = S \\ \\ 0 & \text{if } x \notin T \text{ or } T \not\subseteq S \end{cases}$$

and $\sum\limits_{T \subseteq S} \Gamma_x(T \mid S) = 1$.

For any given DM, $\Gamma_x(T \mid S)$ gives the probability that set T is considered when S is available and the DM starts at option x. However, if the DM follows the specific network consideration procedure explored in this model, these consideration probabilities will have more structure and will be explicitly related to the distribution F and network of connections g. To this end, I define a random network consideration set mapping, denoted $G_x(T \mid S)$. For ease of notation, let $\mathcal{G}_T^S = \{g \in \mathcal{G} \mid g[S] \text{ is T-Connected}\}$. The definition of a random network consideration set mapping is as follows:

Definition 3. *Given a distribution F on $\mathcal{G}$, a **random network consideration set mapping** is a function $G_x : \Omega x \Omega \rightarrow [0,1]$ such that the following is true:*

$$G_x(T \mid S) = \begin{cases} 1 & \textit{if } \{x\} = T = S \\ \sum_{g \in \mathcal{G}_T^S} f(g) & \textit{if } \{x\} \subseteq T \subseteq S \\ 0 & \textit{otherwise} \end{cases} \tag{1.4}$$

The extreme cases, where $\{x\} = T = S$, where $x \notin T$, or where $T \nsubseteq S$ are trivial. For the non-trivial case, a random network consideration set mapping can be thought of as being constructed according to a sequential process. First, given S, restrict attention of networks to $g[S]$. This is done to include those networks in $G_x(T \mid S)$ that are not T-Connected only due to some element $t' \in X \setminus S$. Second, among all $g[S] \in \mathcal{G}$, consider those that are T-Connected, further restricting attention to $\mathcal{G}_T^S \subseteq \mathcal{G}$. Finally, given these networks that connect set T under available set S, the probability that T is considered is simply the sum of the probabilities of each network occurring.

1.5.2 Necessary Properties

We first look at a natural implication of the definition of T-Connectedness for some network $g[S]$. Consider both $g[S]$ and $g[S \cup \{a\}]$, for $a \notin S$. A network $g[S]$ that is T-Connected for some $T \subseteq S$ may or may not stay T-Connected for the same set under $S \cup \{a\}$. The new element a may be connected to some $t \in T$ or it may not. What is certain, however, is that all of the elements in t remained connected to one another when this new element is added. This is formally stated in the following Lemma. All proofs for this section are contained in the Appendix.

Lemma 1. *For any g such that $g[S]$ is T-Connected for some $T \subseteq S$, $g[S']$ is T'-Connected for some unique T' such that $T \subseteq T'$ and $T' \subseteq S'$, for all $S \subseteq S'$. Equivalently, $\mathcal{G}_T^S \subseteq \bigcup_{T' \subseteq S' : T \subseteq T'} \mathcal{G}_{T'}^{S'}$ for all $S \subseteq S'$.*

This leads us to our first characteristic of random network consideration set mappings:[10]

A1. RNC Upward Monotonicity For each $x \in T \subseteq S$, $\Gamma_x(T \mid S) \leq \sum_{T' \subseteq S' : T \subseteq T'} \Gamma_x(T' \mid S')$

That random network consideration set mappings should satisfy this condition should be obvious. If a network with attention restricted to S leads to set T being considered (i.e. it is part of the sum that makes up $G_x(T \mid S)$), then by Lemma 1, that same network is T'-Connected for some $T' \subseteq S'$ with $S \subseteq S'$. Then that same network will appear as part of the sum that makes up some (unique) $G_x(T' \mid S')$. In short, if a network is included on the left-hand side of A1, it will show up on the right-hand side as well. To see why this expression does not hold with equality, consider the following example network:

Then when $S = \{x, y, z\}$ is available, the set $T = \{x, y, z\}$ is considered with probability $f(g_1)$ when the starting point is x. However, when $S' = \{w, x, y, z\}$ is available, the probability that T' is considered is $f(g_1) + f(g_2)$. Under g_2, node z is

[10]Note that each of these properties of RNC will be written for general starting-point contingent consideration set mappings, denoted $\Gamma_x(T \mid S)$. Random network consideration set mappings ($G_x(T \mid S)$) are one particular instance of starting-point contingent random consideration set mappings and they satisfy the general properties proposed in this section.

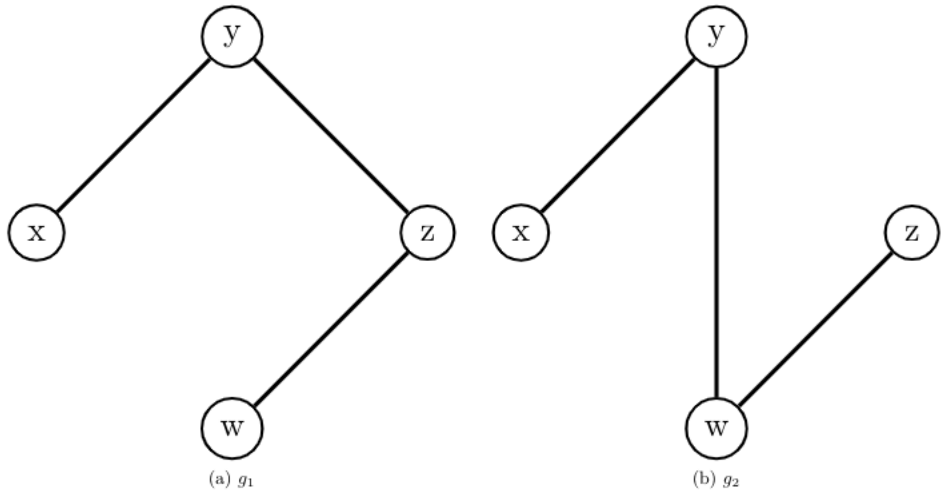


Figure 1.7: Monotonicity Example

connected to x and y through node w. When w is removed, z cannot connect to x or y under g_2, resulting in $f(g_2)$ being included on the right-hand side of the RNC Upward Monotonicity axiom, but not the left hand side when $T = \{x, y, z\}$. Notice that, in this example, had we considered $T = \{x, y\}$, the expression would have held with equality.

This property is clearly a stochastic generalization of the Upward Monotonicity property of Masatlioglu and Suleymanov (2017). When we restrict attention to $\Gamma_x(T \mid S) \in \{0, 1\}$, A1 is equivalent to B1. This, along with the relationship between other properties of RNC and NC, will be further explored in the proof to Proposition 1.

I now consider what the effect of changing the starting point might have on the probability of a given subset being considered. First, it should be clear that comparing $\Gamma_x(T \mid S)$ to $\Gamma_y(T \mid S)$ for some $y \notin T$ is essentially trivial. If $x \in T$, $y \notin T$ will imply that T cannot be considered from y, and so $\Gamma_x(T \mid S) \geq \Gamma_y(T \mid S) = 0$.

But this is not informative, since $\Gamma_x(T \mid S) \geq 0$ by definition. However, when switching from x to y while both are in T, we reveal a fundamental characteristic of random network consideration:

A2. RNC Symmetry: $\Gamma_x(T|S) = \Gamma_y(T|S)$ for all $\{x, y\} \subseteq T \subseteq S$

This comes from the straightforward observation that $\mathcal{G}_T^S$ does not depend on the starting point and will be the same for all $t \in T$. Thus, in the non-trivial case in Definition 3, which is satisfied by both x and y (since $\{x, y\} \subseteq T$), we are summing over the same set of networks, resulting in the same consideration probabilities for each $t \in T$ as the starting point.

In a similar fashion to the RNC Upward Monotonicity property above, RNC Symmetry is a generalization of NC Symmetry in the deterministic case. In the Symmetry property for NC, the inclusion of y in $\Gamma_x(S)$ indicates that y is connected to x when S is available. When we change the starting point to y, this connection remains. Contrary to A2, Symmetry in NC restricts the DM to follow the same sub-network of consideration on S in both extended decision problems (x, S) and (y, S), such that not only x, but the entirety of $\Gamma_x(S)$ must be considered in (y, S), since we know that y and x exist on the same sub-network. In RNC, it is not required that <u>the same</u> sub-network be followed by the DM to construct the consideration set in both extended decision problems. The only requirement is that the probability of a given sub-network occurring does not depend on the starting point, conditional on the starting points being included in that sub-network.

Finally, we explore what this random network structure implies about the connectedness of certain options. Consider the following: there exists some $T \subseteq S$ with $z \in T$ where $\Gamma_x(T \mid S) > 0$. This then implies that there exists some network that connects x to z when S is available which occurs with positive probability. Now consider the removal of some element $y \in S$. If there still exists some $T' \subseteq S \setminus \{y\}$ with $z \in T'$ and $\Gamma_x(T' \mid S \setminus \{y\}) > 0$, we do not learn anything additional about how x, y, and z are connected, other than that y is not required for there to exist a path between z and x. However, if there exists no such T', we learn that y <u>is required</u> for there to be a path between x and z. This information illuminates direct relationships between x and y and between y and z and it leads us to our final necessary property of random network consideration set mappings:

A3. RNC Path Connectedness If $\exists T \subseteq S$ with $z \in T$ such that $\Gamma_x(T \mid S) > 0$ and $\nexists T' \subseteq S \setminus \{y\}$ such that $z \in T'$ and $\Gamma_x(T' \mid S \setminus \{y\}) > 0$, then the following must hold:

(a) $\exists T'' \subseteq S \setminus \{z\}$ with $y \in T''$, such that $\Gamma_x(T'' \mid S \setminus \{z\}) > 0$

(b) $\exists T''' \subseteq S \setminus \{x\}$ with $z \in T'''$, such that $\Gamma_y(T''' \mid S \setminus \{x\}) > 0$

The hypothesis, as mentioned previously, reveals that y is required to establish a connection between z and x. In other words, all such paths that have x and z as terminal nodes must include y as an intermediate node. Then we can break up one such path into its x-y and y-z sub-paths. The x-y sub-path survives when z is removed, which means y is considered with some positive probability when z is removed (the first implication of the above). Similarly, the y-z sub-path survives when x is removed, so z is considered with some positive probability when x is removed and the starting point is y.

Again, RNC Path Connectedness is a stochastic generalization of Path Connectedness in the deterministic NC model. Thus far, I have claimed that each of A1 - A3 are stochastic generalizations of consideration set properties of NC. The following Proposition captures this notion, with the proof contained in the Appendix.

Proposition 1. *If $\Gamma_x(T \mid S)$ is a random consideration set mapping such that i) $\Gamma_x(T \mid S) \in \{0,1\}$ for all $x \in T \subseteq S$ and ii) Γ_x satisfies A1 - A3, then $\Gamma_x(S)$ satisfies B1 - B3 where $\Gamma_x(S) = T$ for $\Gamma_x(T \mid S) = 1$.*

Finally, it should be clear at this point that RNC consideration set mappings necessarily exhibit all of the above properties. Proposition 2 captures this idea.

Proposition 2. *If a random consideration set mapping has a random network consideration set mapping representation, it satisfies RNC Symmetry, RNC Upward Monotonicity, and RNC Path Connectedness.*

Proof. Suppose a random consideration set mapping Γ has a random network consideration set mapping G:

RNC Symmetry Consider any $T \subset S$ with $\{x, y\} \subseteq T$ and $x \neq y$. If $G_x(T \mid S) > 0$, then there exists some network $g[S]$ that is T-Connected. Since $y \in T$, $g[S]$ will also be included in $G_y(T \mid S)$. Therefore, $G_x(T \mid S) \leq G_y(T \mid S)$. $G_y(T \mid S) \leq G_x(T \mid S)$ by the same logic. Finally, if $G_x(T \mid S) = 0$, then there are no $y \in \mathcal{G}$ such that $g[S]$ is T-Connected. This will hold regardless of the starting point in T, so $G_y(T \mid S) = 0$ as well. Then RNC Symmetry holds

RNC Upward Monotonicity Given Lemma 1, the proof is trivial. With $\}[\mathcal{S}]_{\mathcal{T}} \subseteq \bigcup_{T' \subseteq S' : T \subseteq T'} \mathcal{G}_{T'}^{S'}$ for all $S \subseteq S'$, the statement follows directly from the definition of $G_x(T \mid S)$.

RNC Path Connectedness Let T be such that $z \in T$ with $G_x(T \mid S) > 0$. Then there exists some $g[S] \in \mathcal{G}_T^S$ where $f(g[S]) > 0$. Since $\nexists T' \subseteq S \setminus \{y\}$ with $z \in T'$ such that $G_x(T' \mid S \setminus \{y\}) > 0$, then every path that connects x to z under $g[S]$ must include y as an intermediate node. To see why this is the case, consider some x-z path in $g[S]$ that does not include y as an intermediate node. When y is removed from S, this path remains (since y was not on this path under $g[S]$) and $G_x(T' \mid S \setminus \{y\}) > 0$ for $T' = \{j \mid j \text{ is connected to some node on this } x\text{-}z \text{ path in } g[S]\}$ since $f(g[S]) > 0$. Since there exists an x-y-z path in y^S, we can consider each sub-path independently.

Consider the x-y sub-path. When z is removed from S, this path survives, and if we let $T'' = \{j \in S \setminus \{z\} \mid j \text{ is connected to some node on the } x\text{-}z \text{ path in } g[S]\}$, then $G_x(T'' \mid S \setminus \{z\}) > 0$, since $f(g[S]) > 0$.

By similar logic, $G_y(T''' \mid S \setminus \{x\}) > 0$ for $T''' = \{j \in S \setminus \{x\} \mid j \text{ is connected to some node on t}$

$\square$

1.5.3 Choice Rule

The DM is also endowed with an antisymmetric and transitive preference relation, $\succ$. Given the consideration set $T \subseteq S$, after the realization of the random network process, the DM chooses the $\succ -$maximal element of T. We thus define a

Random Network Choice (or, abusing abbreviations, RNC) as follows:

Definition 4. *A choice rule π is a* ***random network choice (RNC)*** *if there exists a set of networks $\mathcal{G}$ on X, a probability distribution over these networks F, and an antisymmetric, transitive preference relation $\succ$ on X, such that:*

$$\pi_y(x \mid S) = \sum_{\{x,y\} \subseteq T \subseteq S} \mathbb{1}\{x \text{ is } \succ\text{-best in } T\} G_y(T \mid S) \tag{1.5}$$

where G is a random network consideration set mapping.

1.5.4 Revealed Preference

In general, it may be possible for there to be multiple RNC representations of a given π. Suppressing notation for X and $\mathcal{G}$, we denote a given RNC representation using only the distribution over networks and the preference relation, $(F, \succ)$. Given some choice rule π, we denote the set of possible RNC representations as $(F^\pi, \succ^\pi$ $) = \{(F^1, \succ^1), ..., (F^N, \succ^N)\}$. Following Masatlioglu et al. (2012) and Lleras et al. (2017) and erring on the side of being conservative in assessing revealed preference, we say that "x is revealed preferred to y" (denoted $x \succ y$) if $x \succ^i y$ for all $\succ^i$ such that $(F^i, \succ^i) \in (F^\pi, \succ^\pi)$.

With the possibility of multiple RNC representations for a given π, under what conditions can we guarantee that $x \succ y$ for each representation? It turns out that a very simple condition captures all aspects of revealed preference.

Lemma 2. *For any RNC π, x is revealed preferred to y if $\exists S \subseteq X$ such that:*

$$\pi_y(x \mid S) > 0 \tag{1.6}$$

Given this method to reconstruct $\succ$ for an RNC π, we then ask whether this condition is sufficient to reveal preferences completely. We first define the following binary relations to assist in exploring this idea:

Definition 5. *Let P be a binary relation X^2 such that xPy if $\exists S \subseteq X$ such that:*

$$\pi_y(x \mid S) > 0 \tag{1.7}$$

Further, let P_R be the transitive closure of P on X^2.

We utilize this binary relation to obtain the following helpful result:

Lemma 3. *For some RNC π, x is revealed preferred to y if, and only if, xP_Ry.*

1.5.4.1 Connection to RAM

As mentioned previously, the random network consideration set mappings of RNC satisfy a starting-point contingent version of the monotonicity condition laid out in Cattaneo et al. (2017). We call this condition "Starting-Point Monotonicity" and it is as follows:

$$\text{Starting-Point Monotonicity: } \Gamma_x(T \mid S) \leq \Gamma_x(T \mid S \setminus \{a\}) \tag{1.8}$$

for $a \notin T$. It should be clear that, provided Γ_x has an RNC representation, it will satisfy this Starting-Point Monotonicity condition. If an element a is removed from S that was not in T, we can now include networks in $\Gamma_x(T \mid S \setminus \{a\})$ that were not T-Connected exclusively because of the element a. Cattaneo et al. (2017) show, albeit with no starting-point contingent attention, that if some random choice rule has a RAM representation, preferences can be revealed in the following manner. Under RAM, x is revealed preferred to y if, and only if, the following holds:

$$\exists S \text{ such that } \pi_z(x \mid S) > \pi_z(x \mid S \setminus \{y\}) \tag{1.9}$$

One might surmise that since RNC satisfies Starting-Point Monotonicity, that preferences could also be revealed using condition 1.9. In this case, we may be missing some revealed preferences by only considering P_R as defined above. Lemma 4 shows that this worry is unfounded: under RNC, if x and y satisfy condition 1.9, $(x, y) \in P$ as defined above.

Lemma 4. *Let π be an RNC and let x and y be such that there exists some set $S \supseteq \{x, y, z\}$ such that the following holds:*

$$\pi_z(x \mid S) > \pi_z(x \mid S \setminus \{y\}) \tag{1.10}$$

Then $(x, y) \in P$ and x is revealed preferred to y.

1.5.5 Pseudo-Markovian RNC

As mentioned in Section 1.5.4, it is possible that, for a given set of consideration set or choice data consistent with RNC, there may be multiple $(F, \succ)$ representations thereof. Consider the following example. The choice probabilities in Table 1.11 come from an RNC where all possible networks on three options have the same probability of $\frac{1}{8}$ and $\succ$ is such that $x \succ y \succ z$. Given this choice data, one could work in the opposite direction, uniquely identifying probabilities of a subset of networks that are consistent with an RNC representation. As it turns out, in this example, one cannot uniquely identify probabilities for a subset of these networks given choice probabilities alone.

	Available Set S			
	$\{x, y\}$	$\{x, z\}$	$\{y, z\}$	$\{x, y, z\}$
$\pi_z(x \mid S)$	-	$\frac{1}{2}$	0	$\frac{5}{8}$
$\pi_z(y \mid S)$	-	0	$\frac{1}{2}$	$\frac{1}{8}$
$\pi_z(z \mid S)$	-	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$
$\pi_y(x \mid S)$	$\frac{1}{2}$	-	0	$\frac{5}{8}$
$\pi_y(y \mid S)$	$\frac{1}{2}$	-	1	$\frac{3}{8}$

Table 1.11: Choice Data for $f(g) = \frac{1}{8}$

To see this starkly, first consider the RNC representation of this choice data that actually generated these choice probabilities (i.e. F such that $f(g) = \frac{1}{8}$ for all g). Now, consider those networks under which z is isolated or only connected to x, as displayed in Figure 1.8.

It is easy to see that, for any $\epsilon \in [0, \frac{1}{8}]$, a newly constructed pair $(F', \succ)$ will also represent the data in Table 1.11, where F' is constructed as follows:

$$f'(g) = \begin{cases} \frac{1}{8} & \text{if } g \notin \{g_1, g_2, g_3, g_4\} \\ \frac{1}{8} + \epsilon & \text{if } g \in \{g_1, g_3\} \\ \frac{1}{8} - \epsilon & \text{if } g \in \{g_2, g_4\} \end{cases} \tag{1.11}$$

with g_1, g_2, g_3, and g_4 as in the figure above. In words, F' is simply F adjusted by ϵ for some of the networks. Thus, even for this very simple case, the most general RNC model can lead to an infinite number of plausible representations of some given π. However, this may not be a desirable property in some empirical environments. In this section, I consider a stochastic special case of RNC, Pseudo-Markovian RNC (PM-RNC, hereafter) that does end up exhibiting a unique representation for some π.

In this special case, we consider only RNC representations of a particular form, where we add restrictions on F in the following manner:

Definition 6. *An RNC $(F, \succ)$ is a **Pseudo-Markovian RNC (PM-RNC)** if there exists a matrix μ with entries $\mu_{ij} \in [0, 1]$ where for each network $g \in \mathcal{G}$, the probability of g occurring, $f(g)$ can be written as follows:*

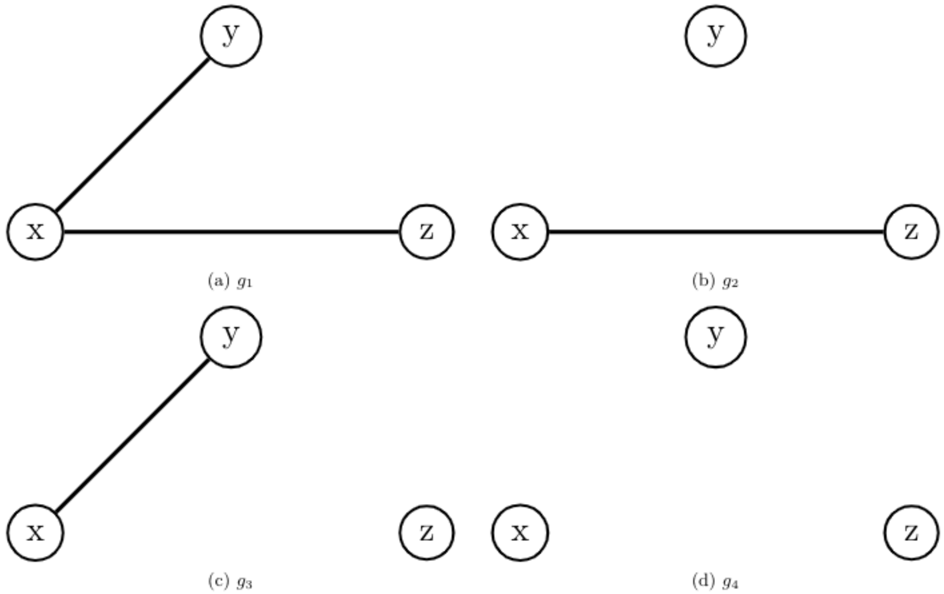


Figure 1.8: Networks under which z is isolated or only connected to x

$$f(g) = \prod_{(i,j)\in X^2} [\mathbb{1}\{g_{ij} = 1\}\mu_{ij} + \mathbb{1}\{g_{ij} = 0\}(1 - \mu_{ij})] \tag{1.12}$$

For PM-RNC, we use the notation $M_x(T \mid S)$ to refer to the RNC G_x under this particular formulation. Further, it suffices to represent the entire distribution over networks as a weighted network, with the weight on the connection between options i and j as μ_{ij}. We denote a PM-RNC representation as $(\mu, \succ)$.

A benefit of considering the special case of PM-RNC is that the distribution over networks in each PM-RNC representation is unique:

Lemma 5. *Let π be a PM-RNC with representations $(\mu^\pi, \succ^\pi)$. Then $\mu^i = \bar{\mu}$ for all $(\mu^i, \succ^i)$ representations of π (i.e. μ is unique).*

We can build up intuition for this form of a starting point contingent random

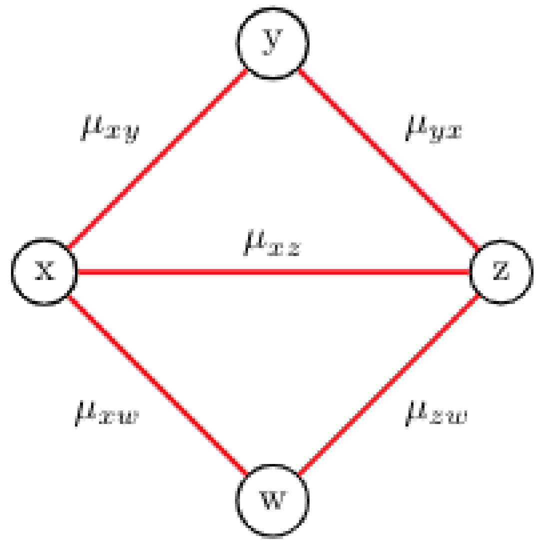


Figure 1.9: Network g with Four Nodes

consideration set mapping by working through the following example:[11]

Suppose that we are interested in computing $G_x(T \mid S)$ under the network in Figure 1.9. Consider $G_x(\{x, y, z\} \mid \{w, x, y, z\})$. The DM starts at option x, considering node x with probability 1. Consideration of the set $\{x, y, z\}$ from the set $\{w, x, y, z\}$ can then follow any of the T-Connected networks under S, shown in Figure 1.10.

[11]Note that, in all figures, $\mu_{ij} = 0$ is represented as the absence of a connection between nodes i and j.

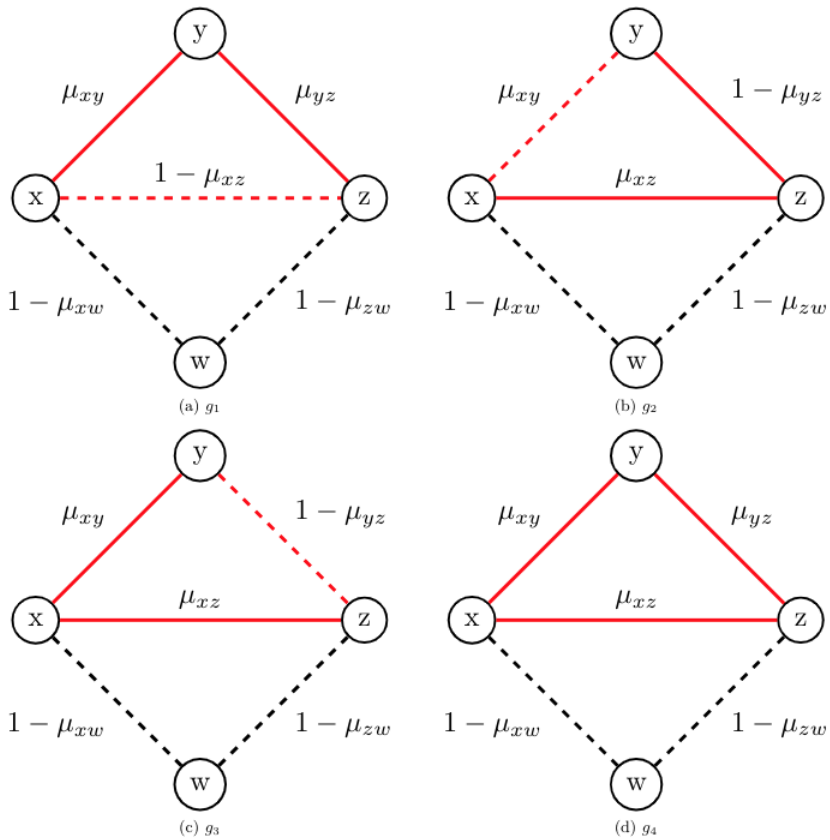


Figure 1.10: Networks that generate Consideration Set $\{x, y, z\}$

In each case, we consider the probability that consideration spills over from node x to nodes y and z, but not to node w (hence its inclusion in each of the networks in Figure 1.10). Taking g_1 as an example, we then construct $f(g_1)$ by multiplying together μ_{xy} and μ_{yz} (consideration spills over from node x to y, then from y to z), then $(1-\mu_{xz})$ (consideration does not spillover from x to z or vice versa), and finally by $(1-\mu_{xw})$ and $(1-\mu_{zw})$ (consideration does not spillover from x or z to w).

We can then calculate the probability of the consideration set being $\{x, y, z\}$ as follows:

$$\begin{aligned}
G_x(\{x,y,z\} \mid \{x,y,z,w\}) = & \quad f(g_1) + f(g_2) + f(g_3) + f(g_4) \\
= & \quad \mu_{xy}\mu_{yz} \cdot (1-\mu_{xz}) \cdot (1-\mu_{xw})(1-\mu_{zw}) \\
& + \mu_{xz}\mu_{yz} \cdot (1-\mu_{xy}) \cdot (1-\mu_{xw})(1-\mu_{zw}) \\
& + \mu_{xy}\mu_{xz} \cdot (1-\mu_{yz}) \cdot (1-\mu_{xw})(1-\mu_{zw}) \\
& + \mu_{xy}\mu_{xz}\mu_{yz} \cdot (1-\mu_{xw})(1-\mu_{zw}) \qquad (1.13)
\end{aligned}$$

We can thus use the above procedure to calculate $G_x(T \mid \{x,y,z,w\})$ for any T. But what happens when we add an element to the grand set of alternatives? Now consider the following example:

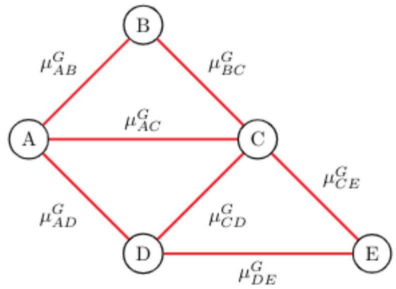


Figure 1.11: Network g' with Five Nodes

When inspecting the same consideration set $T = \{x, y, z\}$ under $\{x, y, z, w, v\}$, we have to include the possibility that consideration now spills over to node v from node z. In this case, we use the following constrained sub-networks of G':

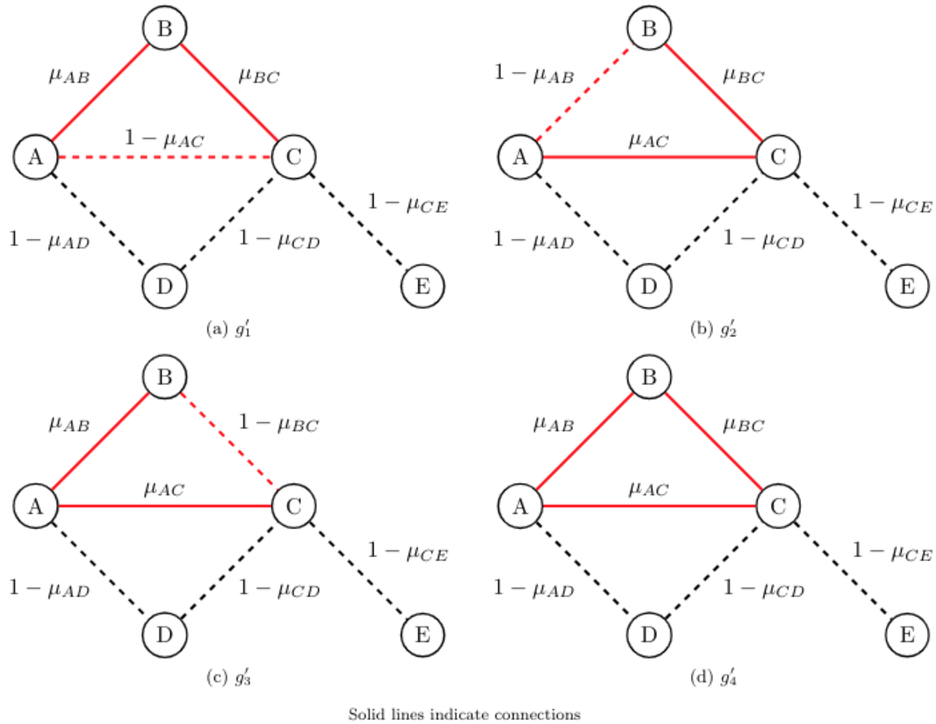


Figure 1.12: Networks that generate Consideration Set $\{A, B, C\}$

Looking just at nodes $\{x, y, z\}$ and connection weights between them, we can see that this is identical to the networks in Figure 1.10. This is straightforward, as nothing has changed in connection weights within the set $\{x, y, z\}$. Additionally, in each network displayed in Figure 1.12, we have to account for the possibility that consideration spills over to node w, just as we did for the networks in Figure 1.10. However, we now have to consider the possibility that node v is considered once z is considered. For this reason, we can write $G_x(\{x, y, z\} \mid \{x, y, z, w, v\})$ as follows:

$$\begin{aligned}
G_x(\{x,y,z\} \mid \{x,y,z,w,v\}) &= f(g_1') + f(g_2') + f(g_3') + f(g_4') \\
&= (1-\mu_{zv})[f(g_1) + f(g_2) + f(g_3) + f(g_4)] \\
&= (1-\mu_{zv})G_x(\{x,y,z\} \mid \{x,y,z,w\}) \qquad (1.14)
\end{aligned}$$

Then as we inspect the same potential consideration set T in larger and larger sets, the probability of set T being considered aggregates by including the probability that none of the added alternatives are considered, accounting for all links between the added alternatives and the set T. Notice that in the above, we don't consider any connection weights between alternatives w and v in any of the networks used to construct $G_x(\{x,y,z\} \mid \{x,y,z,w,v\})$. This is intuitive: since we aren't considering cases where alternative w is considered, we never need to account for the possibility that consideration spills over from w to v (even though there is a positive probability of this happening under the network g', represented by μ_{wv}).

From the above examples, one can intuit an additional property of PM-RNC consideration set mappings beyond those of the general RNC case. The networks enumerated in Figures 1.10 and 1.12 should illuminate the fact that RNC Symmetry holds under PM-RNC. However, a comparison between the examples in Figures 1.9 and 1.11 suggests a more strict form of monotonicity than that required in RNC Upward Monotonicity. This characteristic is given below:

A1. PM-RNC Binary Separability $\Gamma_x(T \mid S \cup \{S'\}) = \Gamma_x(T \mid S) \prod_{z \in S' \setminus S} \prod_{t \in T} \Gamma_z(\{z\} \mid$

$\{t, z\})$,

Proposition 3. *If an RNC G_x has a PM-RNC representation, it satisfies PM-RNC Binary Separability.*

1.6 Results: Random Network Choice

In this section, tests of the more general stochastic properties of RNC are conducted. For each test, in order to generate consideration probabilities, observations are aggregated over all subject, treating each observation as if it came from a representative subject who encountered each problem multiple times. Thus, for a given extended decision problem (x, S) and for some consideration set $T \subseteq S$, $\Gamma_x(T \mid S)$ was set to be equal to the frequency of consideration set T observed in the full data set, conditional on the extended decision problem being (x, S).

1.6.1 RNC Monotonicity

For each T observed with strictly positive probability, RNC Monotoncity is constructed by comparing $\Gamma_x(T)$ to the sum of probabilities over supersets of T for some presented superset of S, offering a direct test of the RNC Monotonicity property. Table 1.12 presents the aggregate mean violations of RNC Monotonicity. Many consideration sets T are feasible for a given (x, S) extended decision problem, in that they are such that on $T \subseteq S$, but they do not occur with positive probability. Then RNC Monotonicity will, by default, be satisfied trivially. While these observations are technically consistent with RNC Monotonicity, they are excluded in the column labeled “NT”, for “Non-Trivial” in Table 1.12. In the aggregate, 12.5% of all observations result in a violation of RNC Monotonicity, compared to 79.8% when testing against the NC model. Even when only considering “Non-Trivial” observations, the rate of Monotonicity violations is considerably lower under RNC than under NC at 39.7% (Wilcoxon signed-rank $p < 0.01$ for $H_0 : \mu = 0.8$).

Table 1.12: Aggregate Test of RNC Monotonicity

	All	NT
Mean	0.125	0.397
Std Error	0.000	0.000
N	9974	3132

Wilcoxon signed-rank $p < 0.01$ for aggregate test of $H_0 : \mu = 0$ for both All and NT
Mann-Whitney $p < 0.01$ for $H_0 : \mu_{All} = \mu_{NT}$
NT results exclude observations where $\Gamma_x(T \mid S) = 0$

Considering the presence of RNC Monotonicity violations even when “trivial” observations are included, I further investigate the determinants of these violations. For a given (x, S) and T pairing, one can imagine several measures as generalizations of the “Distance” measure used in Section 1.4.3.1. First, conditional on S, larger consideration sets T leave less room for supersets to be included in the right-hand side of the RNC Monotonicity expression. Larger sets T more closely approach full consideration of the set S, leaving less room for non-trivial observations of supersets of T under $S' \supseteq S$. Thus, we may expect that violations are more likely to occur as $S \setminus T$ increases in size across observations. Second, when comparing to some set $S' \supseteq S$, the size of $S' \setminus T$ relative to T may have a similar effect. These factors are considered in Table 1.13. Additional options in $S \setminus T$ actually increase the likelihood of a violation occurring by 0.299 percentage points each, though this effect is only marginally significant. Consistent with the hypothesis presented above, each additional option in $S' \setminus T$ decreases the likelihood of a violation occurring by 2.57 percentage points each. This result is further confirmed in Figure 1.13, where average RNC monotonicity violations are plotted by quartile of $|S' \setminus T|$. This leads to an interesting comparative result relative to NC. Recall that in Section 1.4.3.1, it is shown that monotonicity violations were ubiquitous once the distance between S and S' became sufficiently large. Here, the opposite is true: holding the size of T constant, as options are added to S', RNC Monotonicity violations become less common.

Table 1.13: Determinants of RNC Monotonicity Violations

	Model 1
$\|S \setminus T\|$	0.00299*
	(0.00175)
$\|S' \setminus T\|$	-0.0257***
	(0.00185)
Observations	9974

Standard errors in parentheses
Marginal effects from logistic regression specifications
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

While none of these results on the determinants of RNC Monotonicity Violations are implied by the RNC model directly - clearly the RNC model implies no violations of RNC Monotonicity at all - I view the analysis above as reasonable starting points for further generalizations of network consideration in the stochastic case. It is possible that generalizing RNC further to include adaptive consideration behavior, directed random network structures, and/or other features may directly imply the results above.

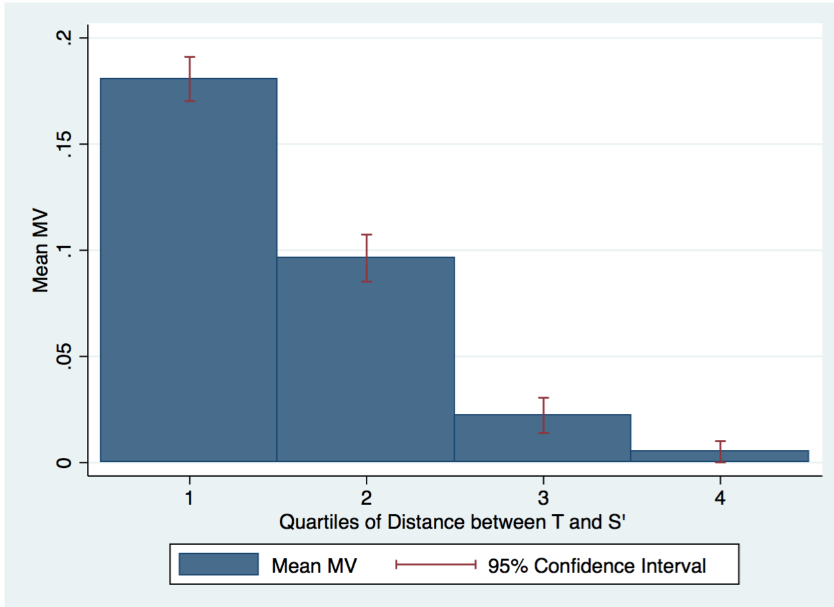


Figure 1.13: RNC Monotonicity Violations by $|S' \setminus T|$ Quartile

1.6.2 RNC Symmetry

While the subject-level data reveals a number of violations of the deterministic Symmetry property in NC, it may be possible that aggregate consideration set probabilities are consistent with RNC. This is tested using two methodologies: first, by conditional logit regression estimation and second, by individual difference-in-means tests for each pair of consideration set and available set.

Initially, each observation used for the conditional logistic regression specifications consists of a subject, an extended decision problem (i.e. a starting point that is an element of $\{x, y\}$ and an available set A_i), and the set of options considered by the subject (i.e. the observed consideration set). A test of RNC Symmetry consists of a test of whether the probability of a given consideration set being observed is dependent on the starting point in $\{x, y\}$, given that the consideration set contains both starting points. To this end, for each of the five available sets used for the Symmetry extended decision problems, I treat each observed unique consideration set that includes both starting points as an "available consideration set" that the subject may then "choose". Thus, a case for the purposes of these conditional logistic regressions is defined as a subject - available set pair, with the unique consideration sets which include both starting points observed in the data for that available set across all subjects constituting the "available consideration sets" from which this subject can "choose." Note that, for a given case, each of these consideration sets is offered to the subject as an available consideration set twice: once for each extended decision problem that utilized the available set for this case. This is done to allow for the possibility that the same consideration set was chosen by an individual subject in both extended decision problems that utilize the available set for this case. Thus, in these conditional logit specifications, the dependent variable Choose indicates whether the "available consideration set" was "chosen" for an individual case. The lone dependent variable, Starting Point, is a binary variable that takes the value 1 when the starting point for the observation is y and 0 otherwise.

Of the 553 distinct consideration sets observed for the 10 extended decision problems constructed to test the Symmetry property, 235 were such that $\{x, y\} \subseteq T$. Results at the aggregate level are displayed in Table 1.14. When we aggregate over all possible available sets ($N \in \{5, 10, 15, 20, 25\}$), we see that there is no relationship between the starting point and whether a consideration set is "chosen." This result is robust to whether a separate conditional logit regression is run on each available set individually, as can be see in Table 1.15. There are thus broad, early indications that consideration set formation is consistent with RNC Symmetry.

Table 1.14: RNC Symmetry: Aggregate

	All
Choose	
Starting Point	-0.0862
	(0.0848)
Observations	38082

Standard errors in parentheses

Odds ratios from conditional logit regression specifications

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.15: RNC Symmetry Regressions

	N = 5	N = 10	N = 15	N = 20	N = 25
Choose					
Starting Point	-0.252	-0.0155	-0.155	-0.0216	0.134
	(0.168)	(0.176)	(0.186)	(0.208)	(0.232)
Observations	776	5100	13448	11088	7670

Standard errors in parentheses

Odds ratios from conditional logit regression specifications

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

While aggregate results support RNC Symmetry according to these conditional logistic regression models, it is possible that additional information provided to the subject in the Context environment might have an effect on RNC Symmetry, especially in light of the slight, though statistically insignificant, difference in symmetry violations at the individual level between the Baseline and Context environments. The same regression specification used in Table 1.15 and is conducted separately for each environment in Tables 1.16 and 1.17. This reveals a significant, though mixed, effect of the starting point in certain treatment - N combinations.

Table 1.16: RNC Symmetry Regressions: Baseline

	N = 5	N = 10	N = 15	N = 20	N = 25
Choose					
Starting Point	-0.319	-0.358	-0.407*	-0.437	-0.443
	(0.232)	(0.229)	(0.236)	(0.274)	(0.302)
Observations	400	3000	8036	6006	4602

Standard errors in parentheses

Odds ratios from conditional logit regression specifications

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.17: RNC Symmetry Regressions: Context

	N = 5	N = 10	N = 15	N = 20	N = 25
Choose					
Starting Point	-0.178	0.527*	0.288	0.614*	1.147***
	(0.244)	(0.291)	(0.312)	(0.345)	(0.434)
Observations	376	2100	5412	5082	3068

Standard errors in parentheses

Odds ratios from conditional logit regression specifications

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Each consideration set observed in the data individually can be examined directly by looking at $\Gamma_x(T \mid S)$ and $\Gamma_y(T \mid S)$ for each $T \subseteq S$ combination. If the frequency of T conditional on S being available is significantly different between Γ_x and Γ_y, this is a violation of RNC Symmetry. We thus conduct a Wilcoxon sign-rank test on each of the Y consideration sets where $\{x, y\} \subset T$. Note, however, that if a given T is only ever chosen once in the entire sample, a sign-rank test will result in an insignificant difference by starting point.[12] A relatively conservative approach is used here, where only those consideration sets that occur with non-trivial frequency, defined as "having occurred more than once across all 10 symmetry extended decision problems," are included in this analysis. Of the 235 distinct consideration sets observed in these extended problems that satisfy $\{x, y, \} \in T$, 183

[12]Consideration set - available set pair has 214 observations, 107 for each starting point. If the consideration set only occurs once, the sign-rank test will result in $p > 0.10$.

occurred only once. The remaining consideration sets are then used to construct $\Gamma_x(T \mid S)$ and $\Gamma_y(T \mid S)$ for each S such that T appeared at least once across (x, S) and (y, S). This resulted in 69 separate tests of $T \subset S$ pairs.[13] Of these 69 tests, only 5 resulted in a statistically significant rejection of $H_0 : \Gamma_x(T \mid S) = \Gamma_y(T \mid S)$, only 7.23%. We therefore find robust support for RNC Symmetry regardless of the test method (conditional logit vs. rank-sum at the consideration set - available set pair level).

1.6.3 RNC Path Connectedness

Finally, RNC Path Connectedness is tested by systematically considering subsets of the experimental data according to whether they satisfy the hypotheses of RNC Path Connectedness for each potential case. Recall that, in the construction of the extended decision problems used to test Path Connectedness in NC, four different cases resulted from varying the option used in each extended decision problem. Table 1.18 presents the results of aggregate tests of RNC Path Connectedness separately for each case.

Recall that under the RNC Path Connectedness, the hypothesis of this property would be endogenously determined by consideration set data in the experiment:

$$\exists T \subseteq S \text{ with } z \in T \text{ such that } \Gamma_x(T \mid S) > 0 \text{ and } \nexists T' \subseteq S \setminus \{y\} \text{ such that } z \in T'$$
$$\text{and } \Gamma_x(T' \mid S \setminus \{y\}) > 0,$$

In order to structure tests of RNC Path Connectedness, I then consider, for each

[13]Note that this is greater than $235 - 183 = 52$. This is due to the fact that several consideration sets were observed under multiple available sets.

case, the largest number of observations such that this hypothesis is satisfied when aggregating over subjects to construct each $\Gamma_x(T \mid S)$. Then N in Table 1.18 can be interpreted as the number of subjects (out of 107) who satisfy the particular hypothesis of the case under consideration. Cases 2 and 4 are clearly less stringent tests of RNC Path Connectedness; all 107 subjects satisfy the hypothesis of these cases.

The implication portion of the RNC Path Connectedness property is framed around "the existence" of some set that includes z (y) and is considered with positive probability. This is equivalent to a positive frequency of $z(y)$ being observed for the observations considered in each case. Then for each case, the fact that $Prob(Y)$ and $Prob(Z)$ are positive in Table 1.18 indicates that the dataset, as a whole, is consistent with this property. This should not be a surprise, even considering the fact that there are non-trivial violations of other properties of RNC consideration sets documented in this section. RNC Path Connectedness, like Path Connectedness, is a relatively weak requirement to impose on consideration set probabilities.

Table 1.18: RNC Path Connectedness

	Case 1		Case 2		Case 3		Case 4	
	Prob(Y)	Prob(Z)	Prob(Y)	Prob(Z)	Prob(Y)	Prob(Z)	Prob(Y)	Prob(Z)
Mean	0.327	0.327	0.925	0.561	0.227	0.091	0.879	0.178
Std Error	0.068	0.068	0.026	0.048	0.045	0.031	0.032	0.037
N	49	49	107	107	88	88	107	107

Result 4. *Aggregate consideration set frequencies are largely consistent with RNC:*

- *Only 12.5% of all observations violate RNC Monotonicity*

- *Fewer than 8% of all consideration sets observed in the aggregate data violate*

RNC Symmetry

- *Aggregate results are wholly consistent with RNC Path Connectedness*

1.6.4 PM-RNC Binary Separability

In addition to RNC Monotonicity, RNC Symmetry, and RNC Path Connectedness, consideration set data consistent with PM-RNC necessarily must satisfy the additional Binary Separability property. Results of tests thereof using this experimental data are presented in this section.

Notice that Binary Separability will necessarily result in consideration set probabilities such that $\Gamma_x(T \mid S \cup S')$ is weakly less than $\Gamma_x(T \mid S)$, since PM-RNC simply takes the latter and multiplies it by the product of a number of probabilities between 0 and 1, inclusive. A clear violation of Binary Separability would thus consist of an observation of $\Gamma_x(T \mid S \cup S') > \Gamma_x(T \mid S)$. Aggregate tests of violations of this type, which I term "First-Order Binary Separability Violations," are presented in Table 1.19.

Table 1.19: Aggregate Test of First Order PM-RNC Binary Separability Violations

	All	NT
Mean	0.122	0.282
Std Error	0.003	0.007
N	9974	4308

Wilcoxon signed-rank $p < 0.01$ for aggregate test of $H_0 : \mu = 0$ in each case
NT: $\Gamma_x(T \mid S) = \Gamma_x(T \mid S') = 0$ excluded

In the aggregate, 12.2% of observations constitute first-order Binary Separability violations, with the proportion of such violations jumping to 28.2% when we isolate attention to only "Non-Trivial" observations as defined above.

In order to provide a more finely-tuned test of Binary Separability, the expression provided in the PM-RNC Binary Separability property is constructed for each observation that does not constitute such a first-order Binary Separability Violation. Recall this expression:

$$\Gamma_x(T \mid S \cup S') = \Gamma_x(T \mid S) \prod_{z \in S' \setminus S} \prod_{t \in T} \Gamma_z(\{z\} \mid \{t, z\})$$

In the above, the construction of $\Gamma_x(T \mid S \cup S')$ and $\Gamma_x(T \mid S)$ is straightforward and conducted as in previous analyses in this section. However, the construction of product on the right-hand side, $\prod_{z \in S' \setminus S} \prod_{t \in T} \Gamma_z(\{z\} \mid \{t, z\})$, is not so clear, given that subjects are never presented with an extended decision problem of the form $(z, \{t, z\})$ for any pair of options. In the PM-RNC model, $\Gamma_z(\{z\} \mid \{t, z\})$ is clearly equal to $1 - \mu_{zt}$, or the probability that consideration does not spill over from z to t in the binary comparison. Thus, in order to construct this nested product, I utilize an estimated $1 - \mu_{zt}$ as a proxy for $\Gamma_z(\{z\} \mid \{t, z\})$. For each pair of options (i, j) presented to subjects in the experiment, an observed μ_{ij} is estimated by calculating the frequency with which subjects navigate from i to j, or from j to i, conditional on a link being provided in the experimental interface. Note that this frequency is across all observed links between i and j, regardless of the direction followed, so that the resulting estimated weighted network of μ_{ij} is undirected. Only those estimated μ_{ij} which were statistically significantly greater than 0 at the $\alpha = 0.10$ level were taken as non-zero frequencies.[14]

Table 1.20: Aggregate Test of Binary Separability Violations

	All	NT
Mean	0.349	0.989
Std Error	0.005	0.002
N	8759	3093

Wilcoxon signed-rank $p < 0.01$ for aggregate test of $H_0 : \mu = 0$ in each case
NT: $\Gamma_x(T \mid S) = \Gamma_x(T \mid S') = 0$ excluded

Mean direct Binary Separability violations are reported in Table 1.20. Results in Table 1.20 are presented only for those observations that did not constitute a first-order Binary Separability violation. Across all such observations, 34.9% violate Binary Separability directly along with a staggering 98.9% of Non-Trivial observations (as defined above). The latter result is not particularly surprising, since a "violation" as defined in this section does not allow for additional noise in consideration probabilities beyond that directly implied by the strong form of PM-RNC: unless the realized left-hand side of the Binary Separability expression was exactly equal to the estimated right-hand side, the observation was coded as a Binary Separability violation. A more thorough investigation of Binary Separability would thus necessitate analyzing the size of errors in estimation.

Table 1.21 reports the summary statistics of two measures of errors in the estimation of these Binary Separability expressions, compared to the distribution of positive $\Gamma_x(T \mid S \cup S')$ for reference. In an abuse of notation for the sake of brevity, in Table 1.21 and in this discussion, let Γ refer to $\Gamma_x(T \mid S \cup S')$ and $\hat{\Gamma}$ refer to the estimate of $\Gamma_x(T \mid S) \prod_{z \in S' \setminus S} \prod_{t \in T} \Gamma_z(\{z\} \mid \{t, z\})$. Then $\frac{\Gamma - \hat{\Gamma}}{\Gamma}$ gives the normalized

[14]Each μ_{ij} was tested using a Wilcoxon signed-rank test.

difference between the two expressions, conditional on positive Γ, and $\hat{\Gamma} \mid \Gamma = 0$ gives the estimated right-hand side expression, conditional on Γ being equal to zero. Both of these measures of Binary Separability errors are presented only for those observations where a violation is observed.

Table 1.21: Binary Separability Error Size Summary Statistics

	Γ	$\frac{\Gamma-\hat{\Gamma}}{\Gamma}$	$\hat{\Gamma} \mid \Gamma = 0$
Mean	.011	.031	.005
SD	.008	.759	.004
N	1427	178	2881
p1	.009	-3.591	.000
p25	.009	-.143	.002
p50	.009	.253	.005
p75	.009	.471	.007
p99	.047	.938	.016

Notes:
Γ s.t. $\Gamma_x(T \mid S \cup S') > 0$

First compare the first two columns for Γ and $\frac{\Gamma-\hat{\Gamma}}{\Gamma}$. Note that the interquartile range deflation rate of Γ (given by $\frac{\Gamma-\hat{\Gamma}}{\Gamma}$ for p50) is -0.143 to 0.471 implying that there is considerable spread in the PM-RNC Binary Separability estimate of Γ. This implies that PM-RNC Binary Separability is likely too strong an assumption for the given experimental data. However, the distribution of $\hat{\Gamma} \mid \Gamma = 0$ appears to closely resemble that of Γ suggesting that $\hat{\Gamma}$ may yet still present a reasonable estimate of Γ for out-of-sample consideration sets; it's possible that because the experimental data includes a considerable number of observations of $\Gamma = 0$, that PM-RNC Binary Separability is failed because of data limitations. There are thus only mixed, and generally negative, results concerning the fit of PM-RNC Binary Separability to the experimental data set. I view further exploration of PM-RNC Binary Separability using a large field dataset to be a worthwhile avenue for future research on this topic.

1.7 Discussion

The NC and RNC models can be applied to many settings, but in this section I focus on one particular application of note: empirical studies of the effects of advertising. As mentioned previously, Shapiro (2018),Sahni (2016), and Lewis and Nguyen (2015) report evidence of so-called "positive spillovers in advertising," where attention spills over from an advertised good to similar competitor goods. A common refrain in studies such as these is that the positive externalities resulting from these attention spillovers lead to an under-allocation of advertising in the competitive equilibrium. Indeed, Shapiro (2018) presents a supply-side model and accompanying estimates which support this claim.

However, an implicit assumption in the models commonly assumed in the above studies is that these positive attention spillovers affect competing goods at the category level. When this framework is rooted in an RNC model setup, it can be easily seen that a more general treatment of these spillovers as coming from a generalized random network structure will result in more ambiguous welfare implications.

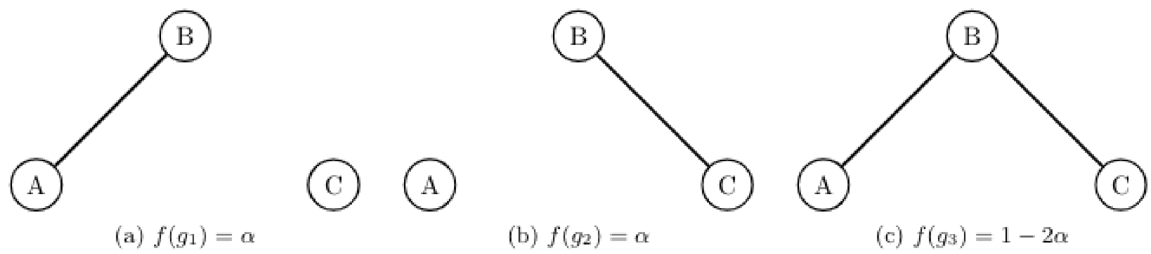


Figure 1.14: An example RNC network

1.7.1 A Simple Advertising Model

Consider a simple RNC Advertising Game with firms A, B, and C, each producing a single good (which I will also denote A, B, and C, respectively), and each choosing whether advertise or not, such that the strategy for each firm is $\sigma_i \in \{0, 1\}$. Advertising ($\sigma_i = 1$) involves a fixed cost of $c > 0$, which is identical across firms. There is a continuum of DMs of mass 1, each of whom exhibits limited stochastic attention according to the RNC model. The distribution over possible networks on $X = \{A, B, C\}$ is given in Figure 1.14, with a restriction on α such that $\alpha \in [0, \frac{1}{2}]$.

The DM's choice of any firm leads to the same level of utility, so I assume that a DM who considers more than one firm is equally likely to choose any of the firms they consider.[15] Further, since RNC is starting-point dependent, I assume that DMs set their starting points based on the firms that are advertising. If more than one firm advertises, the mass of DMs is divided equally among all those firms who advertise. The sequence of the game then works as follows:

[15] Note that this random choice procedure is indeed a departure from the choice procedure used in RNC. This is done for tractability in this particular application and is of no particular consequence to the main point under discussion.

1. Firms simultaneously choose whether to advertise or not
2. DMs observe who advertises and sets a firm as their starting point with equal probability among all those who advertise
   (a) If no firm advertises, DMs consider no firm and choose nothing (resulting in 0 profit for all firms)
3. DMs form stochastic consideration set mappings according to their starting points and the distribution over networks on the set of firms
4. DMs choose each firm that they consider with equal probability
5. Firms realize profit, which is simply the mass of DMs who choose their firm minus the fixed cost of advertising (if the firm chose to advertise)

For a given network g, let $N_i(g)$ represent the set of firms connected to firm i under g by some path (i.e. firm i's "neighbors"). Given a strategy profile σ, let the set of firms who advertise under σ be equal to σ_a. Then for a given strategy profile and distribution over networks on $\{A, B, C\}$, firm i's expected profit is equal to the following:

$$\pi_i = \sum_{j=1}^{3} f(g_j) \frac{|\sigma_a \cap N_i(g_j)|}{|\sigma_a| \cdot |N_i(g_j)|} - c \cdot \sigma_i \tag{1.15}$$

Given this setup, it is straightforward to show that there exists a non-empty subset of the $\alpha - c$ parameter space that lead to i) $\sigma^* = (1, 0, 1)$ as a Nash Equilibrium and ii) $\sigma' = (0, 1, 0)$ not as a Nash Equilibrium, where profit is strictly higher under

σ'. This is captured in the following proposition, with the proof included in the Appendix:

Proposition 4. *In the RNC Advertising Game with parameters α, β, and c, such that (α, β, c) is in the unit cube, there exist non-empty subsets of the parameter space where:*

1. $\sigma^* = (1, 0, 1)$ *is supported as a Nash Equilibrium*

2. $\sigma' = (0, 1, 0)$ *is not supported as a Nash Equilibrium*

3. *Aggregate profit under σ' is higher than under σ^**

This proposition then tells us that the welfare properties of advertising in such an environment will depend on the structure of distribution in the RNC limited consideration on the part of the consumers. In particular, the "category" spillover approach implicitly assumed in previous work in the marketing body of literature is nested in the RNC Advertising Game where $\alpha = 0$. When this is true, advertising will never be undertaken by multiple firms simultaneously in equilibrium. Indeed, for sufficiently high costs ($c > \frac{1}{3}$), no firm will advertise. This is a stark example of the free-riding effect documented in Shapiro (2018), but is only implied by a narrow interpretation of attention spillovers as spillovers occurring equally across an entire category of goods. In the more general case described here, positive attention spillovers actually result in an over-allocation of advertising relative to an alternative allocation with the same levels of revenue.

1.8 Conclusion

In this work, I present a the results of an experiment designed to test for the validity of the deterministic Network Choice (NC) model of Masatlioglu and Suleymanov (2017). Overall inconsistency of the data with NC led to a proposed stochastic generalization thereof in RNC and PM-RNC. These models have applications in the realms of choice architecture, web platform search optimization, and advertising. In the latter case, I view a significant contribution of this work as illuminating a possible model for positive spillovers in advertising behavior, which highlights the need to examine product network structures empirically, especially for the purposes of welfare analysis.

It should be noted that RNC is only one version of a stochastic generalization of NC. Indeed, Cattaneo et al. (2017) present another version in which the network is deterministic and starting points are stochastically determined. While the experimental data presented herein was predominantly consistent with RNC, there remain non-trivial violations thereof to be investigated in either further generalizations of RNC or other attempts at modelling stochastic network consideration. This is likely a fruitful avenue for future research.

Additionally, the mixed results of the PM-RNC model, combined with the RNC results, suggest that there is a nested model of network consideration with more structure than RNC (which potentially leads to a unique representation), but less than that of PM-RNC. Future theoretical research can likely shed light on the structure and validity of models between the two.

Chapter 2: The Relevance of Irrelevant Information

Co-authored with Emel Filiz-Ozbay and Erkut Ozbay

2.1 Introduction

In many decision problems, unavailable options along with irrelevant attributes are presented to decision makers. For example, a search on Amazon.com for televisions yields 1,239 different alternatives, 753 of which are unavailable at the time of search.[1] Additionally, these televisions are described by a great number of attributes: e.g. Refresh Rates, backlighting vs. no backlighting, size dimensions, availability of Wi-Fi connectivity, SMART vs non-SMART functions, number and types of inputs, etc. Many of these attributes may be irrelevant to some decision makers.

Consider some additional examples of unavailable alternatives: In a restaurant menu, unavailable items may still be listed in the menu with a sold out note. A health insurance buyer will go over the insurance plans, some of which she is not qualified to purchase. A local event ticket website may list events that are sold-out. Also, consider some more examples of irrelevant attributes: Insurance coverage for care

[1] Site accessed 02/02/2017.

related to pregnancy may be presented to someone who could never get pregnant. The US Food and Drug Administration requires standardized nutrition label on food and beverage packages including fat, cholesterol, protein, and carbohydrate even when they are 0%, such as for a bottled water. Smartphones will list available service providers, even though this set will not vary across available smartphones.[2] From the perspective of classical rational choice theory, decision makers have free disposal of irrelevant information: they can costlessly ignore unavailable options and irrelevant attributes, and hence the presentation of such irrelevant information would not lead to different choices than those made when it is not presented. We experimentally demonstrate that the presentation set matters, providing evidence that the free disposal of irrelevant information is a non-trivial assumption in many contexts.

Our experiment is designed to test the effects of presenting irrelevant information in two dimensions. In a differentiated product setting, the decision problems presented to subjects vary according to a) the presentation of options in a set of alternatives that can never be chosen (hereinafter referred to as unavailable options) and b) the presentation of attributes that have no value (i.e. that enter into a linear utility function with an attribute-level coefficient of zero; hereinafter referred to as irrelevant attributes). We find significant evidence that the presence of both unavailable options and irrelevant attributes increases the frequency of sub-optimal choice, but that adding one without the other (i.e. unavailable options with no

[2]An attribute that does not vary across available options may be utility relevant, but it is certainly not decision relevant information in that it does not meaningfully distinguish one good from another.

irrelevant attributes or irrelevant attributes with no unavailable options) does not.

Furthermore, motivated by the variation in online shopping websites allowing consumers to sort on the products based on the attributes they consider relevant, as well as allowing them to exclude the unavailable alternatives, we ask if individuals are willing to pay to reduce the amount of irrelevant information presented to them. We show that subjects are willing to pay significant positive amounts not to see unavailable alternatives or irrelevant information. Such a payment is mainly due to the reduction in mistakes and time costs caused by the presence of unavailable options and irrelevant attributes. Nevertheless, individuals may have a "preference for simplicity" in the presentation of information implying an additional cost, a cognitive cost of ignoring the irrelevant information. In order to identify such a cognitive cost, we analyze the willingness to pay (WTP) of the subjects who always chose optimally and who experience no additional time costs in the presence of unavailable options and irrelevant attributes. Our results indicate that even these subjects are willing to pay positive amounts to change the presentation set.

To our knowledge, unavailable alternatives have only been studied in the context of the decoy effect, which is the presentation of an alternative that increases the preference for a target alternative. Although in a typical experiment on decoys, the decoy alternative is available in the choice set, Soltani et al. (2012) showed that displaying an inferior good during an evaluation stage, but making it unavailable at the selection stage, also generates the decoy effect. Also, the phantom decoy alternatives that are superior to another target option, but unavailable at the time of choice, increases the preference for the inferior target option (see e.g. Farquhar and Pratkanis (1993)). The crucial difference between the decoy effect experiments and our experiment is that in our setup the unavailable alternative does not create a reference point for another alternative, hence it allows us to directly investigate the impact of the presentation set.

Our experiment also complements the experimental literature investigating the effects of relevant information on choice optimality. In particular, Caplin et al. (2011) find that additional (available) options and increased "complexity" (additional relevant attributes in our context) lead to increased mistake rates. Also, Reutskaja et al. (2011) present evidence from an eye-tracking experiment that subjects are unable to optimize over an entire set (given a large enough alternative set), but can optimize quite well over a subset (see also Gabaix et al. (2006)). One contribution of our work herein is to show that a similar effect is present for adding unavailable alternatives and increasing the number of irrelevant attributes.

Finally, it is worth mentioning that existing bounded rationality models that are capable of explaining the sub-optimal decisions build on the *available* alternatives and the *relevant* attributes. For example, in the limited consideration models, the DM creates a "consideration set" from the *available* set of alternatives and then chooses from the maximal element of the "consideration set" according to some rational preference relation (see e.g. Masatlioglu et al. (2012), Manzini and Mariotti (2007; 2012; 2014), and Lleras et al. (2017)). Also, according to the boundedly rational model that focuses on attributes, the salience theory of choice, certain *relevant* attributes may appear to be "more salient" to a DM than others, causing them to be overweighted in the decision-making process (see Bordalo et al. (2012), Bordalo et al. (2013), and Bordalo et al. (2016)). Our results highlight that the DM considers not only the alternative set and the relevant attributes but also the presentation set in which unavailable options and the irrelevant attributes are presented. The presentation of a decision problem can be viewed as a "frame" as in Salant and Rubinstein (2008). However, if the DM chooses the best option when the presentation set is simple, but chooses a subotimal option by using a boundedly rational model, such as a model of satisficing as in Simon (1955), when the presentation set is more complex, such an extended choice function induces a choice correspondence that cannot be described as the maximization of a transitive, binary relation. We discuss this formally in Section 2.5.

The rest of the paper is organized as follows. Section 2.2 explains the design of the experiments in detail. Section 2.3 presents the experimental results. Section 2.5 discusses our results in light of extant theory and suggests a "presentation set" approach to modelling choice and Section 2.6 concludes.

2.2 Experimental Procedure

The experiments were run at the Experimental Economics Lab at the University of Maryland (EEL-UMD). All participants were undergraduate students at the University of Maryland. The data was collected in 14 sessions and there were two parts in each session. No subject participated in more than one session. Sessions lasted about 90 minutes each. The subjects answered forty decision problems in Part 1, and a subject's willingness to pay to eliminate unavailable options and irrelevant attributes were elicited in Part 2. In each session the subjects were asked to sign a consent form first and then they were given written experimental instructions (provided in Appendix C.9) which were also read to them by the experimenter. The instructions for Part 2 were given after Part 1 of the experiment was completed.

The experiment is programmed in z-Tree (Fischbacher, 2007). All amounts in the experiment were denominated in Experimental Currency Units (ECU). The final earnings of a subject was the sum of her payoffs in ten randomly selected decision problems (out of forty) in Part 1, her payoffs in two decision problems she answered in Part 2, the outcome of the Becker et al. (1964) (BDM) mechanism in Part 2, and the participation fee of $7. The payoffs in the experiment were converted to US dollars at the conversion rate of 10 ECU = 1 USD. Cash payments were made at the conclusion of the experiment in private. The average payments were $27.90 (including a $7 participation fee).

Each decision problem in the experiment asked the subjects to choose from five available options and each option had five relevant attributes. Each attribute of an option was an integer from {1,2,..., 9} and it could be negative or positive. The value of an option for a subject was the sum of its attributes.[3] The subjects knew that their payoff from a decision problem would be the value of their chosen option if that decision problem was selected for payment at the end of the experiment. Figure 2.1 provides an example of both an available option and an unavailable option presented to the subjects (see Appendix C.9 for examples of the decision screen presented to

[3]A similar design wherein the value of an option is the sum of its displayed attributes is used in Caplin et al. (2011).

subjects in each decision problem). Note that the header of each column indicates whether an attribute enters to the option value as a positive or negative integer (plus or minus sign). Whether a column should be added, subtracted, or ignored when calculating the value of an option was only indicated in this header row, so this information had to be continually referenced as the subject considered options at lower positions on the screen. In some decision problems, some of the attributes did not enter the value of an option and those were indicated by zero at the header.[4] In Figure 2.1, there are ten attributes with a zero in the header and this means that the option had ten irrelevant attributes which did not affect the value of the option for the subjects. In a given decision problem, there were either five relevant attributes (each one with either positive or negative integer value from $\{1, 2, \ldots, 9\}$) or fifteen attributes where five of them were relevant and ten of them were irrelevant. The value of an option was the sum of its positive and negative attributes and it was a randomly generated positive number to guarantee that the subjects will not lose money by choosing an option.

Figure 2.1: Options with 5 Relevant and 10 Irrelevant Attributes

		+	+	**0**	**0**	**0**	+	**0**	**0**	**0**	**0**	+	**0**	**0**	**-**	**0**
☐	Option 1	three	four	three	one	seven	four	four	two	six	two	eight	five	two	six	one
	Option 2	one	eight	two	six	one	five	nine	two	six	two	eight	three	one	seven	nine

Regardless of the type of decision problem, the matrix of information presented to the subject took up the entire screen. This design was chosen to abstract away from possible confounds that lie in the <u>way</u> that information is presented. No matter

[4]Our design of varying irrelevant information in two dimensions will later be shown to create symmetric difficulty for subjects. Even though one may think that the perceptual operations required to solve a task are very different in these two dimensions (keeping track of payoffs horizontally and vertically), the impact of these two dimensions on decision makers turn out to be similar.

which type of decision problem the subject faced, their eyes were forced to scan the entirety of the screen in order to fully process all relevant information. In this way we abstract away from the possibility that subjects are more capable of processing less (or more) visual space on a computer screen.

In each decision problem, the subjects needed to choose one of the five available options in 75 seconds.[5] In some decision problems they were presented fifteen options and told that only five of them were available to choose from. The other ten were shown on their screens but the subjects were not allowed to choose any of those. O_iA_j is the notation for a decision problem with i options and j attributes. The decision problems that were used in the experiment had $i, j \in \{5, 15\}$; in each case the effective numbers of options and attributes were five, i.e. if the number of options or attributes on a screen was fifteen, then ten of those were either unavailable options or irrelevant (zero) attributes. The order of the decision problems were randomized at the session-individual level (i.e. Subject 1, for instance, in each session, saw the same order of decision problems; with 16 subjects per session, we therefore have 16 distinct decision problem orderings).

Once Part 1 of the experiment was completed, subjects received instructions for Part 2. The aim of Part 2 was to elicit subjects' willingness to pay to eliminate unavailable options or irrelevant attributes to estimate the cost of ignoring irrelevant information. A BDM mechanism was used to measure subjects willingness to pay to remove irrelevant information in one direction. Hence, we elicited the subjects' WTP in four different directions: moving from i) $O_{15}A_5 \rightarrow O_5A_5$, ii)

[5]Subjects earned a payoff of $0 if they didn't make a choice within 75 seconds.

$O_5A_{15} \to O_5A_5$, iii) $O_{15}A_{15} \to O_5A_{15}$, and iv) $O_{15}A_{15} \to O_{15}A_5$. The distribution of selling prices used in the BDM procedure (and explained to subjects) was uniform from 0 to 15 ECU. These four BDM elicitation procedures were conducted across two treatments for Part 2 of our experiment: a "low information" treatment and a "high information" treatment. Seven sessions were conducted for each treatment. In the "low information" treatment, BDM procedures were run for (i) and (ii) - WTP was elicited for removal of options or attributes, given that irrelevant information in the opposite dimension was not present. In "high information" treatments, BDM procedures were run for (iii) and (iv) - WTP was elicited for removal of options or attributes, given that irrelevant information in the opposite dimension was present and cannot be eliminated. Hence, we elicited the cost of ignoring 10 unavailable options and cost of ignoring 10 irrelevant attributes separately and in two different informational environments. Note that a given subject completed two BDM procedures, with roughly half of our subjects completing (i) and (ii) and half of them completing (iii) and (iv). We chose this between-subject design to eliminate a possible framing effect where a subject may have thought that she was expected to price the elimination of unavailable options or irrelevant alternatives differently depending on the amount of information in the other dimension. Table 2.1 summarizes the treatments of the experiment.

Subjects completed Parts 1 and 2 without being provided any feedback on their performance in earlier decision problems similar to the experiments in related literature. First, we did not provide feedback after each decision problem in Part 1 in order to avoid any reference dependence or triggering new emotions such as re-

Table 2.1: Treatment Summary

Treatment	# of Sessions	# of Subjects	Part 1: Decisions	Part 2: BDM
Low Info	7	112	40 Decisions	$O_{15}A_5 \to O_5A_5$ and $O_5A_{15} \to O_5A_5$
High Info	7	110	40 Decisions	$O_{15}A_{15} \to O_5A_{15}$ and $O_{15}A_{15} \to O_{15}A_5$

gret. For example, a subject may work harder than otherwise she would if she knows that she would receive feedback on how suboptimal her decision was. Second, we do not provide aggregate feedback at the end of Part 1 to avoid unnecessary priming and to more closely approximate an analogous real-world setting. Direct feedback regarding mistake rates and/or time spent in each decision problem type may induce the subject to think that they should be willing to pay to eliminate irrelevant information, even if the subject does not intrinsically possess such a preference. We view the potential effect of feedback in this setting as analogous to an experimenter demand effect.

After the completion of Parts 1 and 2, the subjects answered a demographic questionnaire where they reported gender, age, college major, self-reported GPA, SAT, and ACT scores, and they were given the chance to explain their decisions in Part 2 of the experiment.

2.3 Experimental Results

Our main hypothesis is that unavailable options and irrelevant attributes cause cognitive overload for the decision makers and this leads to sub-optimal choice. In the following analysis, we say that a "mistake" has been made in an individual decision problem when the subject failed to select the highest valued available option presented within the time limit of 75 seconds. If no option was chosen, this is treated as a "timeout", but not as a mistake. When timeouts are treated as mistakes, results are qualitatively similar.

2.3.1 Part 1: Decision Task

In this section we present the results from Part 1 of the experiment. We begin with aggregate results and then investigate individual-level heterogeneity and learning effects.

2.3.1.1 Aggregate Results

Table C.25 presents the mistake rate for each type of decision problem O_iA_j in the aggregate data for $i, j \in \{5, 15\}$, treating timeouts as mistakes, calculating the "mistake rate" for each treatment as the average of subject-level mistake rate. Note that the addition of unavailable options and irrelevant attributes alone does not generate significantly larger mistake rates relative to the benchmark O_5A_5 (p-values 0.584 and 0.653, respectively for decision problem types $O_{15}A_5$ and O_5A_{15}). However, conditional on the presence of either unavailable options or irrelevant attributes (in types $O_{15}A_5$ and O_5A_{15}), the addition of irrelevant information in the opposite dimension does increase mistake rates by about 50% (p-value 0.000 in each case). Thus, in the aggregate, both unavailable options and irrelevant attributes are necessary to generate increased mistake rates. We believe that this is evidence that our design does not favor one type of irrelevant information over the other. If, for some reason, our design explicitly allowed for easier processing of either unavailable options or irrelevant attributes, we'd expect to see that mistake rates would respond to an increase in irrelevant information in only one dimension. This is clearly not the case. As such, we'd expect our results to be robust to permutations of our design, for example, where the matrix of displayed data was transposed. The results are qualitatively similar when we do not count timeouts as mistakes and these can be found in Appendix C.11.1. Additionally, when we instead measure welfare loss from sub-optimal choice by i) the rank of the chosen option among the available options or ii) normalized loss in ECU relative to optimal choice, our main result survives. These analyses can be found in Appendix C.11.4.

Table 2.2: Mistake Rates: Timeouts as Mistakes

		O_5	O_{15}
A_5	Mean	0.213	0.218
	Std Error	0.013	0.013
	N	222	222
A_{15}	Mean	0.228	0.337
	Std Error	0.012	0.016
	N	222	222

$p = 0.000$ for $O_{15}A_5 \rightarrow O_{15}A_{15}$, $O_5A_{15} \rightarrow O_{15}A_{15}$, and $O_5A_5 \rightarrow O_{15}A_{15}$
$p > 0.100$ otherwise.

Note that when a subject finds a decision problem more challenging, she may react to this in two ways: (i) she may take more time to make decision and this may or may not lead to an optimal choice; (ii) she may run out of time and computer may record this as a sub-optimal choice. Even though the mistake rates in Table C.25 do not change much when only the number of options is increased while the number of attributes are kept at 5 (from O_5A_5 to $O_{15}A_5$) and when only the number of attributes is increased while the number of options are kept at 5 (from O_5A_5 to O_5A_{15}), this does not necessarily mean that the subjects find the increased number of options or attributes in only one dimension not challenging. This increase in the difficulty of the decision problem may also appear as increased time required to submit a decision. Table 2.3 reports on the average time (in seconds) at which subjects submit a decision in each type of decision problem. Observations where the subject did not submit a decision in the allotted time were are excluded in Table 2.3 just as they were in Table C.25. For results that treat timeouts as the maximum time allotted (i.e. $time = 75$) and for the sub-sample where the subject chose the correct (optimal) option, see Tables C.26 and C.27 in Appendix C.11.1, respectively; results are not qualitatively different from those presented in Table 2.3.

Note that adding irrelevant information in any dimension (i.e. unavailable options or irrelevant attributes) increases the time spent on each decision problem in Table 2.3. However, this difference is not statistically significant when moving from O_5A_5 to $O_{15}A_5$. Time costs increase much more substantially when irrelevant information in one dimension is already present. For example, the time spent increases by just over one second on average with the addition of unavailable options when there are no irrelevant attributes displayed (in the first row of Table 2.3), but increases by nearly 4 seconds when there are irrelevant attributes displayed (in the second row of Table 2.3). A similar effect is present for the addition of irrelevant attributes. Furthermore, from Table 2.3 we may surmise that irrelevant attributes increase time spent more than unavailable options: time spent increases more on average when moving vertically down in Table 2.3 than when we move horizontally across it. Both these interaction and asymmetry effects will be investigated further in the next subsection.

Table 2.3: Time: No Timeouts

		O_5	O_{15}
A_5	Mean	48.605	49.926
	Std Error	0.712	0.680
	N	222	222
A_{15}	Mean	52.935	56.365
	Std Error	0.780	0.810
	N	222	222

p = 0.00 for $O_5A_5 \rightarrow O_5A_{15}$, $O_{15}A_5 \rightarrow O_{15}A_{15}$, $O_5A_{15} \rightarrow O_{15}A_{15}$, $O_5A_5 \rightarrow O_{15}A_{15}$, and $O_{15}A_5 \rightarrow O_5A_{15}$
$p > 0.10$ for $O_5A_5 \rightarrow O_{15}A_5$

Finally, given that there is a time limit of 75 seconds for each decision problem, the increased difficulty that could arise from the presentation of irrelevant information could also increase the rate at which timeouts occur in each type of decision problem. Recall that subjects earn zero in the case of a timeout and letting 75 seconds pass without a choice is worse than choosing randomly. Timeouts are not prevalent in our data: only 4.67% of decision problems resulted in a timeout. 60.31% of timeouts occurred within the first ten periods; 31.16% occurred in the first period. Further, note that our choice of a time threshold is somewhat arbitrary: we could have easily chosen to give subjects more (or less) time to complete each decision problem. As such, we ignore timeouts as a significant concern for the remainder of our analysis, conducting all tests conditional on experiencing no timeouts.[6]

From all of the above, we are left with the following main aggregate results: i) irrelevant attributes and unavailable options are both necessary to generate increased mistake rates, and ii) time costs are increased by irrelevant information displayed in either dimension. We summarize these findings in Result 5. In order to investigate each of these in more detail, we conduct regression analysis to control for individual-level heterogeneity and learning in the following subsection.

Result 5. *Irrelevant information presented in a decision problem can affect choice using several disparate measures:*

- *Unavailable options and irrelevant attributes jointly generate increased mistake rates.*
- *Both unavailable options and irrelevant attributes generate increased time costs.*

2.3.1.2 Individual Heterogeneity

To investigate subject-level heterogeneity in the mistake rate, we conduct logistic regressions controlling for learning, gender, and academic achievement effects. Table 2.4 reports regression results where the dependent variable is "Mistake" and the independent variables are varied in different models specified. "Mistake" is a binary variable with 1 corresponding to the subject failing to select the element with

[6]There were four subjects who experienced timeouts in more than 20% of their decision problems. They are included in the sample upon which all analysis is conducted, but results are not qualitatively different if they are excluded.

the maximal value in the set of (available) alternatives. It is equal to 0 otherwise. In all models, the independent variables are as follows: "Options" is a dummy variable indicating the presence of 10 additional unavailable options displayed (i.e. Options is equal to 1 for type $O_{15}A_5$ and $O_{15}A_{15}$ decision problems and it is 0 otherwise), "Attributes" is defined analogously for irrelevant attributes (i.e. Attributes = 1 for type O_5A_{15} and $O_{15}A_{15}$ decision problems), "Options * Attributes" is the interaction between the type dummies, "Female" is a dummy variable indicating whether the subject is female, "English" is a dummy variable indicating whether the subject's native language is English, "Economics/Business" is a dummy variable indicated whether the subject's major is in the University of Maryland Economics Department or Business School, and "Period" is the period in which the decision problem was presented. Reported coefficients are calculated marginal effects. Standard errors are clustered at the Subject level.

Cognitive Scores were calculated using a combination of responses on the Demographic Questionnaire. Responses for GPA, SAT, and ACT were normalized as in Cohen et al. (1999) and Filiz-Ozbay et al. (2016): Let j be the variable under consideration with $j \in \{\text{GPA, SAT, ACT}\}$, μ_i^j be the value of variable j for subject i, μ_{max}^j be the maximum value of j in the subject population, and μ_{min}^j be the minimum value of j in the subject population. Then let $\hat{\mu}_i^j$, the normalized value of variable j for subject i, be defined as follows:

$$\hat{\mu}_i^j = \frac{\mu_i^j - \mu_{min}^j}{\mu_{max}^j - \mu_{min}^j}$$

such that $\hat{\mu}_i^j$ can be interpreted as the measure of j for subject i, normalized by the distribution of j in the subject population. Some subjects were missing one or more measures for $j \in \{\text{GPA, SAT, ACT}\}$, since these measures were self-reported (and some subjects could not recall their scores on one or more of these measures). As such, the Cognitive Score for subject i was set to $\hat{\mu}_i^{GPA}$ if the subject reported a feasible GPA, $\hat{\mu}_i^{SAT}$ if a feasible GPA score was missing and the subject reported a feasible SAT score, and $\hat{\mu}_i^{ACT}$ if both a feasible GPA and SAT score was missing and the subject reported a feasible ACT score. GPA Scores were given precedent in the calculation of Cognitive Scores because most subjects could reliably report these while SAT Scores took precedent over ACT Scores because it is more common for University of Maryland, College Park undergraduates to have taken the SAT. Results based on using GPA only are presented in Appendix C.11.3.

In addition to the above specified independent variables, we include two more variables in all models: "Position" and "Positive". Variable "Position" is simply the position, from 1 to 15, of the optimal available option that is displayed. Previous work, including Caplin et al. (2011), has shown that subjects often search a list from top to bottom, implying that optimal options displayed lower-down on the list have a lower probability of being chosen due to the early termination of search. We thus include this variable as a control in each of our model specifications, its coefficient being significant and positive in all instances: subjects make more mistakes and spend more time when the optimal option is presented further down a list of alternatives. Variable "Positive" is the number of positive relevant attributes displayed in the decision problem, ranging from three to five.[7] There are potentially two reasons why "Positive" would matter in a given decision problem: i) a subject responds with increased effort in the presence of stronger incentives and ii) subjects find the task less difficult with fewer subtraction operations. The first comes from the fact that the expected value of the optimal available option is increasing in the number of positive attributes. Subjects may then work harder or stop search later in the presence of five positive attributes than in the presence of, say, three positive attributes. It also may be that subtraction operations are more difficult cognitively than addition operations such that the difficulty of the task is decreasing in the number of positive attributes. Our results are consistent with the latter explanation. The coefficient on "Positive" is negative and significant in all regression specifications.

Finally, any effects of irrelevant information that we may find could possibly be due simply to the increased complexity of the decision problem when irrelevant information is added, not due to the mere presence of irrelevant information. For example, adding unavailable options to a decision problem forces the DM to have to "skip" more visual information on the screen in order to evaluate an individual available option, since whether an attribute is positive or negative is displayed at the top of the screen. Similarly, irrelevant attributes force the DM to interrupt the evaluation process, visually "skip" a column of irrelevant information, and then

[7] Our data generation process gave equal weight to the possibility of having a positive or negative relevant attribute. However, we only used generated decision problems that i) had a unique optimal available option and ii) had all positive-valued available options. Thus, the range of the number of positive available options in the generated dataset is more restrictive than that which would be generated without these constraints.

continue with evaluation. Therefore, we define "Attribute Complexity" and "Option Complexity" as the number of "skips" required for full search/evaluation in the decision problem. For example, Option 1 in the example Figure 2.1 above, has a "Option Complexity" equal to 3 (since there are essentially three groups of irrelevant attributes encountered for full evaluation of the option). In the baseline O_5A_5 decision problems, both of these variables are set equal to 0. When "Options"("Attributes") is equal to 1, "Option Complexity" ("Attribute Complexity") varies between 2 and 5 in the realized data.

The regressions in Table 2.4 are conducted on the sub-sample where the submission is made in under 75 seconds. As mentioned above, specifications that treat timeouts as mistakes are qualitatively similar to those presented here. In Model 1, we replicate the aggregate result that can be seen in Table 2: unavailable options and irrelevant attributes increase the mistake rate when presented jointly. Having irrelevant information in both of these dimensions increases the mistake rate by up to 9.52 percentage points (in Model 4). Moreover, this effect is not due to the "complexity" of the decision problem in the presence of irrelevant information, as both Attribute Complexity and Option Complexity are insignificant in Model 4. We see considerable subject-level heterogeneity. Subjects who have higher Cognitive Scores make fewer mistakes. Women make more mistakes on average: being female increases the mistake rate by up to 9.31 percentage points (in Models 2, 3, and 4). We find no evidence of learning; in both models, the coefficient on "Period" is statistically insignificant.[8]

[8] Results are qualitatively similar if we conduct fixed effect panel regressions for all specifica-

Table 2.4: Mistake Rate Regressions

	Model 1	Model 2	Model 3	Model 4	Timeouts as Mistakes
Options	0.00969	0.00943	-0.0218*	-0.0560**	-0.0733**
	(0.0115)	(0.0115)	(0.0131)	(0.0284)	(0.0292)
Attributes	0.000268	0.000417	-0.00615	-0.0108	0.0172
	(0.0121)	(0.0120)	(0.0122)	(0.0282)	(0.0281)
Options * Attributes	0.0871***	0.0874***	0.0935***	0.0952***	0.100***
	(0.0181)	(0.0180)	(0.0181)	(0.0182)	(0.0182)
Period	0.000314	0.000314	0.000326	0.000323	-0.000895*
	(0.000442)	(0.000438)	(0.000439)	(0.000439)	(0.000461)
Cognitive Score		-0.243***	-0.243***	-0.243***	-0.245***
		(0.0596)	(0.0596)	(0.0596)	(0.0600)
Female		0.0931***	0.0931***	0.0931***	0.0908***
		(0.0218)	(0.0218)	(0.0218)	(0.0222)
Economics/Business		0.00809	0.00793	0.00790	0.0237
		(0.0248)	(0.0248)	(0.0248)	(0.0261)
English		0.00605	0.00599	0.00600	-0.00431
		(0.0234)	(0.0234)	(0.0234)	(0.0247)
Position			0.00433***	0.00486***	0.00572***
			(0.00122)	(0.00126)	(0.00133)
Positive			-0.0277***	-0.0300***	-0.0288***
			(0.00802)	(0.00818)	(0.00835)
Attribute Complexity				0.00120	-0.00172
				(0.00743)	(0.00735)
Option Complexity				0.00929	0.0129*
				(0.00704)	(0.00717)
Observations	8555	8555	8555	8555	8880

Standard errors in parentheses

Marginal effects from logit regression specifications

Robust standard errors reported are clustered at the Subject level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In order to investigate the heterogeneity in time responses to these different types of decisions problems, we present the results of several random-effect Tobit regression models in Table 2.5. Observations are censored below by 0 and above by 75 seconds.[9] In each model presented the dependent variable is Time (measured in seconds), defined as the time at which the subject submits her decision. As in previous model specifications, Models 1 - 4 are conducted on the sub-sample where the time of submission is less than 75 seconds (i.e. excluding timeouts and submissions in the last second). All variables are defined as previously mentioned. In Model 1, we present the simplest model incorporating the effects of the presence of irrelevant information on the time to reach a decision. We find results that are similar to those seen in Table 2.3: irrelevant information displayed in either dimension increases time costs considerably. Further, we confirm that there are interaction effects: that having both unavailable options and irrelevant attributes increases time spent by 1.483 seconds above the individual decision problem type effects. We also discover that irrelevant information has an asymmetric effect on time spent depending on the dimension: irrelevant attributes increase time costs more than unavailable options ($\beta_{Attributes} > \beta_{Options}$; $p - value = 0.000$). Finally, from Model 4 it can be seen that the effect of Options on time to make a decision stems from the increased complexity; Option Complexity is positive and significant in Model 4 while the coefficient on Options is insignificant. This is in keeping with the aggregate results, where we had an insignificant effect of Options in the absence

tions.

[9]To investigate the sensitivity of our results to this choice, we conduct further regressions using lower time thresholds. These can be found in Appendix C.11.2.

of Attributes.

We also find evidence of subject-level heterogeneity. Subjects for whom English is their native language spend less time on average. Female subjects spend less time, but this effect is only marginally significant. There is evidence of learning; "Period" is negative and significant in all model specifications. Note that results are qualitatively similar when our main specification (used in Model 4) is conducted on the sub-sample where the chosen option was the highest valued of the available options (in the model labelled "Correct") or when we treat timeouts as a submission at 75 seconds (in the final model in Table 2.5). Our results are therefore robust to these assumptions.

We summarize all of the aforementioned results in Results 6 and 7:

Result 6. *When controlling for subject-level heterogeneity and learning, we replicate the results found in Result 5. Namely, that irrelevant information can increase the suboptimality of choice and time spent per decision problem.*

Result 7. *We find evidence of subject-level heterogeneity and learning:*

- *There is evidence that female subjects make more mistakes and spend less time on each decision problem.*
- *Subjects with higher Cognitive Scores make fewer mistakes.*
- *There is evidence of learning. Subjects spend less time per decision problem in later periods. However, they do not make fewer mistakes in later periods.*

Table 2.5: Time Regressions

	Model 1	Model 2	Model 3	Model 4	Correct	Timeouts as Time
Options	2.255***	2.267***	1.326***	-1.168	-1.061	-1.778*
	(0.353)	(0.352)	(0.441)	(0.887)	(0.954)	(0.949)
Attributes	5.108***	5.091***	4.807***	3.998***	5.104***	4.920***
	(0.426)	(0.426)	(0.432)	(0.925)	(1.093)	(0.932)
Options * Attributes	1.483***	1.470***	1.733***	1.841***	3.362***	2.122***
	(0.492)	(0.494)	(0.494)	(0.499)	(0.534)	(0.503)
Period	-0.263***	-0.263***	-0.263***	-0.263***	-0.205***	-0.300***
	(0.0256)	(0.0256)	(0.0256)	(0.0255)	(0.0197)	(0.0281)
Cognitive Score		9.507**	9.498**	9.497**	6.191*	8.747**
		(4.101)	(4.101)	(4.102)	(3.342)	(4.276)
Female		-2.567*	-2.569*	-2.569*	-1.337	-2.455*
		(1.356)	(1.356)	(1.356)	(1.121)	(1.364)
Economics/Business		-2.200	-2.205	-2.207	-2.278	-1.479
		(1.565)	(1.565)	(1.565)	(1.398)	(1.601)
English		-3.343**	-3.346**	-3.347**	-2.234*	-3.603**
		(1.433)	(1.432)	(1.432)	(1.332)	(1.434)
Position			0.120***	0.152***	0.194***	0.182***
			(0.0398)	(0.0410)	(0.0456)	(0.0422)
Positive			-1.301***	-1.460***	-1.042***	-1.425***
			(0.264)	(0.269)	(0.281)	(0.281)
Attribute Complexity				0.227	-0.0245	0.120
				(0.227)	(0.295)	(0.228)
Option Complexity				0.692***	0.578**	0.805***
				(0.207)	(0.231)	(0.221)
Observations	8555	8555	8555	8555	6668	8880

Standard errors in parentheses

Marginal effects reported from tobit regressions censored below by 0 and above by 75

Robust standard errors are clustered at the Subject level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.3.2 Part 2: Willingness-To-Pay

Recall that the second part of the experiment elicited subjects WTP to eliminate unavailable options and irrelevant attributes in both "Low Information" and "High Information" environment. Table 2.6 shows the average WTP, measured in Experimental Currency Units (ECUs), for each type of elimination. For reference, recall that the support of the BDM procedure used was $[0, 15]$ ECUs with a uniform distribution.

Table 2.6 can be read from left to right as "WTP to eliminate Attributes given that there are only 5 Options", "WTP to eliminate Options given that there are only 5 Attributes", etc. The first two columns belong to our "Low Information" treatment and the last two belong to our "High Information" treatment. Note that subjects participated in only one of these treatments; a given subject submitted her WTP for either columns 1 and 2 or columns 3 and 4. Thus, when making comparisons between WTP within a particular information treatment (Low or High), we match the data by subject. Let WTP to get rid of information be written as follows: $WTP(X|Y_n)$ where X is the dimension of information they are paying to remove given Y-dimension information with n units. For example, $WTP(A|O_5)$ is the WTP to eliminate 10 irrelevant Attributes, given that five options are present (all of them available). WTP to reduce attributes is significantly higher than WTP to reduce options only in the low information case. (p-value = 0.021 in Wilcoxon Signed-Rank Test with $H_0 : WTP(A \mid O_5) = WTP(O \mid A_5)$).

Tests of whether $WTP(A|O_5)$ is greater (less) than $WTP(A|O_{15})$ and whether

$WTP(O|A_5)$ is greater (less) than $WTP(O|A_{15})$ were conducted un-matched as these were submitted independently by separate subjects. There is no significant difference between WTP to get rid of Attributes or Options by "Low Information" or 'High Information" treatment. Recall that eliminating irrelevant information in one dimension does not affect mistake rates significantly when there is no irrelevant information in the other dimension. However, eliminating irrelevant information in one dimension does affect the mistake rate when there is irrelevant information in both dimensions. Subjects do not seem to anticipate this effect on mistake rates when setting their WTP.

Table 2.6: Willingness to Pay

	Low Information		High Information	
	$WTP(A\|O_5)$	$WTP(O\|A_5)$	$WTP(A\|O_{15})$	$WTP(O\|A_{15})$
Mean	4.473	4.071	4.473	4.373
Std Error	0.286	0.266	0.275	0.273
N	112	112	110	110

$p = 0.021$ for $H_0 : WTP(A|O_5) = WTP(O|A_5)$
$p > 0.100$ otherwise

The regressions reported in Table 2.7 were conducted in order to understand the heterogeneity in the subjects willingness to pay in each of the four directions where irrelevant information could be removed. Table 2.7 displays results aggregated across the Low Information and High Information treatments. Note that in all these regressions, Attributes is a binary variable indicating whether the dependent variable is $WTP(A|O_n)$. When Attributes = 0, the dependent variable is $WTP(O|A_n)$.[10] The variable "High Info" is a dummy variable used to indicate whether the observa-

[10]For these regressions, answers submitted at $time = 75$ seconds are coded as mistakes to avoid collinearity of regressors.

tion is from a High Information treatment. All interaction variables used in Table 2.7 are straightforward.

First we ask if Willingness-To-Pay to eliminate irrelevant information in either dimension is sensitive to measures of performance in Part I of the experiment, despite there being no feedback provided prior to Part II. All three models are Tobit regression specifications with a lower limit of 0 and an upper limit of 15 (i.e. the support of the BDM mechanism used in Part II of the experiment). Note that in all models, Mistakes and Time are a count of the number of mistakes and the sum of time spent across all decision problems in the treatment under consideration for WTP. For example, if a subject in the low information WTP treatment made 7 mistakes across the 10 O_5A_{15} decision problems and spent a total of 500 seconds across these same 10 decision problems, Mistakes would equal 7 and Time would equal 500 for the observation of $WTP(A \mid O_5)$ for this subject.

WTP increases with the incidence of mistakes: Mistakes is positive and significant in all models in Table 2.7. This is somewhat surprising, given that subjects were not provided feedback between Parts I and II of the experiment; it seems that subjects are aware of a general level of optimality of choice and are thus more willing to pay to eliminate irrelevant information if they make more mistakes in the corresponding decision problem type.

Additionally, we ask if these performance measures influence whether WTP is positive: it is possible that WTP itself is not sensitive to individual measures of performance, but that performance in one dimension can affect whether WTP is positive at all. Models 4 through 6 report coefficients from logistic regression specifications where the dependent variable is a binary variable indicating whether WTP is greater than 0. There is only weak evidence that whether WTP is greater than zero is affected by Mistakes and High Information: the coefficients on both of these variables are positive and significant in Model 6 only. Additionally, Model 6 reveals that subjects's WTP may even be less sensitive to Mistakes in the High Information treatment (coefficient on High Info * Mistakes is negative and significant at the $\alpha = 0.05$ level).

Notably, WTP is not sensitive to increased time spent on decision problems in any specification included in Table 2.7. Additionally, subjects are more willing to pay to eliminate irrelevant attributes than unavailable options in the Low Information treatment, but not in the High Information treatment. This is true only at the intensive margin (i.e. in Models 1 and 2) and disappears in Model 3 entirely. We think that (lack of) feedback provided to subjects may prevent them from setting consistent WTP in Low Information and High Information treatments. Further study on the role of feedback in such environments is necessary. We summarize these results in Result 8:

Result 8. *WTP is heterogeneous and sensitive to a number of independent variables:*

- *WTP increases with the number of mistakes made in the relevant decision problem type*
- *There is weak evidence that WTP is higher for Attributes than for Options, but only for the Low Information treatment*
- *Higher mistake rates increase the likelihood that WTP is strictly positive*

Table 2.7: WTP Regressions

	WTP			WTP > 0		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Mistakes	0.198**	0.184**	0.336***	0.205*	0.162	0.441***
	(0.0882)	(0.0886)	(0.113)	(0.108)	(0.101)	(0.146)
Time	-0.00187	-0.00164	0.00116	-0.00177	-0.00141	0.00186
	(0.00180)	(0.00185)	(0.00287)	(0.00187)	(0.00188)	(0.00258)
Attributes	0.311**	0.307**	0.424	0.0817	0.0688	0.147
	(0.149)	(0.149)	(0.271)	(0.144)	(0.146)	(0.311)
High Info	0.0656	0.116	4.971*	-0.167	-0.109	5.759**
	(0.475)	(0.478)	(2.600)	(0.396)	(0.421)	(2.677)
Female		0.0103	-0.133		0.554	0.364
		(0.429)	(0.432)		(0.405)	(0.404)
Cognitive Score		-1.243	-0.997		-1.670*	-1.436
		(1.113)	(1.132)		(0.941)	(1.004)
High Info * Mistakes			-0.325*			-0.441**
			(0.177)			(0.214)
High Info * Time			-0.00591			-0.00731*
			(0.00388)			(0.00388)
High Info * Attributes			-0.350			-0.318
			(0.322)			(0.334)
Constant	4.485***	5.272***	3.187*	2.539**	3.431***	1.006
	(1.092)	(1.265)	(1.733)	(1.252)	(1.330)	(1.627)
sigma						
Constant	3.225***	3.218***	3.197***			
	(0.210)	(0.207)	(0.204)			
Observations	444	444	444	444	444	444

Standard errors in parentheses

Models 1 - 3: Tobit regression specifications with lower limit of 0 and upper limit of 15

Models 4 - 6: Logit regression specifications

Robust standard errors reported are clustered at the Subject level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Robust across these model specifications and treatments is the fact that the constants in these models are always positive and significant (with the exception of the constant in Model 6). For example, consider a subject for whom irrelevant information has no effect: they never make more mistakes when irrelevant information is present and they never spend more (or less) time. This subject would still be willing to pay some amount to eliminate this information. We call this a pure "preference for simplicity" - even in the absence of any effect of irrelevant information on choice, decision makers prefer to exclude it. To our knowledge, ours is the first study to identify such a preference, and this is the "cost of ignoring" in its purest form: there is a preference-based psychological consequence to having to ignore irrelevant information that is not captured by standard measures of the effect of irrelevant information on choice. We investigate this further by analyzing individual WTP for those subjects who experience no increase in mistake rates in the presence of irrelevant information in the following section.

2.3.3 A Preference For Simplicity

To more precisely estimate the extent to which such a preference for simplicity exists, we look at WTP for two categorizations of subjects for a given decision problem: i) those who experience no mistakes and ii) those who make no mistakes and incur no time costs associated with the presence of irrelevant information. Our interpretation of "making no mistakes" differs by the the Informational treatment: for Low Information treatments, a subject is deemed to have made "no mistakes" in decision problems of type O_iA_j if she selected the optimal option in all 10 decision problems of this type; for High Information Treatments, a subject is deemed to have made "no mistakes" in decision problems of type O_iA_j if her mistake rate in O_iA_j was weakly less than her mistake rate in O_iA_{j-10} for $j = 15$ (or $O_{i-10}A_j$, for $i = 15$). In other words, a subject is counted in the first row of Table 2.9 if she indeed made no mistakes for Low Information treatments, or if she made no more mistakes in High Information treatments as a result of irrelevant information in the relevant dimension. For example, a subject in the High Information treatment who made 8 optimal choices in $O_{15}A_5$ and 9 optimal choices in $O_{15}A_{15}$ will be considered to have made "no mistakes" in $O_{15}A_{15}$ because her mistakes didn't increase with the addition of irrelevant attributes. We use two separate interpretations here because using the stricter interpretation (as is used in the Low Information treatments) results in too few subjects satisfying this criteria in the High Treatment for meaningful analysis.

We additionally consider subjects who make no mistakes and incur no additional time costs. A subject is deemed to have incurred no time costs if the difference in the amount of time that she spends in decision problems of type O_iA_j is not significantly different from the amount of time she spends in decision problems of type O_iA_{j-10} for $j = 15$ (or $O_{i-10}A_j$, for $i = 15$). In other words, a subject is counted in the second row of Table 2.9 if she made "no mistakes" as per the interpretation presented in the previous paragraph and she did not spend significantly more time on a type of decision problem as a result of irrelevant information.[11]

[11]In all relevant analysis, "No Mistakes" and "No Mistakes or Time Costs" are defined at the subject-O_iA_j decision problem type level, independent of behavior in other decision problem types. As such, a subject could be considered to have made "No Mistakes" in some decision problems, but not others, and may appear in some cells of Tables 2.8 and 2.9, but not all. These measures

For each sub-group we present the summary statistics of both the WTP level in Table 2.8 and of a dummy variable indicating whether WTP is greater than zero in Table 2.9. The mean WTP and fraction of WTP greater than zero is positive and significant at the 5% level in each case. Additionally, a comparison between the first two rows and the last row of Tables 2.8 and 2.9 reveals that the mean WTP and frequency of positive WTP closely matches that of the overall sample. In fact, only in the $WTP(A \mid O_5)$ case is mean WTP lower for subjects who experience No Mistakes relative to those who do (seen in the left-most cell in the first row of Table 2.8; $p = 0.0859$ in Wilcoxon Signed-Rank Test). Similarly, in Table 2.9, WTP is greater than zero less frequently than for subjects who make mistakes only in the $WTP(O \mid A_5)$ and $WTP(O \mid A_{15})$ cases for subjects who make No Mistakes only ($p = 0.034$, $p = 0.049$ respectively; all other measures in Table 2.9 are not significantly different relative to those for subjects who do make mistakes and/or incur time costs).

Additionally, let $y(I|J_k) = \mathbb{1}\{WTP(I|J_k) > 0\}$ indicate whether WTP to eliminate irrelevant information in the Ith dimension, given that there are k units of information in the Jth dimension, is positive. A Kolmogorov-Smirnov test of equality of distributions fails to reject the null $H_0 : F(y_{\text{mistakes}}(I|J_k)) = F(y_{\text{no mistakes}}(I|J_k))$ for each (I, J_k). Such tests also fail to reject the analogous null for WTP levels themselves ($H_0 : F(WTP_{\text{mistakes}}(I|J_k)) = F(WTP_{\text{no mistakes}}(I|J_k))$).

All of this taken together provides additional evidence that even subjects for whom irrelevant information does not affect the optimality of choice nor increase

do not require any joint conditions over multiple decision problem types for a given subject.

time spent on a decision problem prefer not to see such irrelevant information; there exists a preference for simplicity of the informational environment, even when irrelevant information has no effect on choice. Moreover, a brief look at responses to the open-ended question in our questionnaire reveals similar reasoning for some of our subjects. A subject who made no mistakes responded that "I chose [positive WTP amounts] to relax my eyes a little bit." Another responded that "either one [of eliminating irrelevant attributes or unavailable options] wouldn't be too helpful, but they still kind of help, so I put a low number and if I got it I got it, if I didn't, oh well." One possible explanation for this preference for simplicity may be that there is an additional dimension of cognitive effort spent on these decision problems that is not fully captured by mistake rates or time costs. Said another subject, "[...] unavailable options and attributes are distracting and cause me to work harder and longer when trying to calculate from options and attributes that are actually available. Therefore, I would be willing to pay ECU to get rid of them on the screen in order to work more efficiently and effectively" (emphasis added).

Table 2.8: WTP: No Mistakes

	Low Information		High Information	
	$WTP(A\|O_5)$	$WTP(O\|A_5)$	$WTP(A\|O_{15})$	$WTP(O\|A_{15})$
No Mistakes	3.45	3.2273	4.9167	4.5217
	(.6003)	(.5919)	(.4253)	(.4116)
	20	22	24	23
No Mistakes or Time Costs	3.4444	3.2	5.125	4.6364
	(.6579)	(.6513)	(.5977)	(.4138)
	18	20	16	22
All	4.4732	4.0714	4.4727	4.3727
	(.2856)	(.2663)	(.2748)	(.2727)
	112	112	110	110

Std. Errors in Parentheses

Sample mean > 0 at the $\alpha = 0.05$ level in each instance

Table 2.9: WTP > 0: No Mistakes

	Low Information		High Information	
	$WTP(A\|O_5)$	$WTP(O\|A_5)$	$WTP(A\|O_{15})$	$WTP(O\|A_{15})$
No Mistakes	.85	.7273	.9583	1
	(.0819)	(.0972)	(.0417)	(0)
	20	22	24	23
No Mistakes or Time Costs	.8333	.7	.9375	1
	(.0904)	(.1051)	(.0625)	(0)
	18	20	16	22
All	.8929	.8661	.8636	.8818
	(.0294)	(.0323)	(.0329)	(.0309)
	112	112	110	110

Std. Errors in Parentheses

Sample mean > 0 at the $\alpha = 0.05$ level in each instance

We summarize these results in Result 9:

Result 9. *There is a cost of ignoring irrelevant information that is not measured by mistake rates or time costs: subjects are willing to pay some amount not to see irrelevant information, even when irrelevant information does not affect choice.*

- *When measured by the Constant terms in WTP regressions, this cost is positive.*
- *When measured in an analysis of WTP for subjects who make no additional mistakes in response to irrelevant information, this cost is again positive.*
- *When measured in an analysis of WTP for subjects who make no additional mistakes in response to irrelevant information <u>and</u> spend no additional time in response to irrelevant information, this cost is again positive.*

2.4 Robustness Checks

In order to investigate to what extent our results are sensitive to the design specification used for these tasks, we conducted six additional sessions under alternative designs. Four of these sessions were conducted with alternative designs regarding Part 1 decision tasks and two of these sessions were conducted with alternative designs regarding the Part 2 willingness-to-pay tasks. These sessions are summarized in the following table:

Table 2.10: Robustness Treatment Summary

Treatment	# of Sessions	# of Subjects	Part 1: Decisions	Part 2: BDM
8x8: Low Info	2	32	40 Decisions	$O_8A_5 \rightarrow O_5A_5$ and $O_5A_8 \rightarrow O_5A_5$
8x8: High Info	2	30	40 Decisions	$O_8A_8 \rightarrow O_5A_8$ and $O_8A_8 \rightarrow O_8A_5$
Alt-High Info	2	30	40 Decisions	$O_{15}A_{15} \rightarrow O_5A_5$

In the treatments designated as "8x8" in the above table, decision tasks included a maximum of three unavailable options and three irrelevant attributes relative to the baseline in order to explore the effects of changing the parameter space on our main results. This resulted in decision task treatments O_5A_5, O_5A_8, O_8A_5, and O_8A_8. In the treatment named "Alt-High Info", the decision tasks presented in Part 1 were the same as for the main treatments. However, in Part 2, subjects were asked a single WTP question eliciting WTP to move from $O_{15}A_{15}$ to O_5A_5.

All relevant results are presented in Appendix C.10. In this section, we will highlight several important results that further illuminate the main contributions of this paper.

2.4.1 Convexities in the Mistake Rate Function

In the above analysis, we argue that mistake rates are not affected by unavailable options and irrelevant attributes linearly; both unavailable options and irrelevant attributes are necessary to generate increased mistake rates in our main experimental task. However, an apt reader may notice that with five available options each with five relevant attributes in each treatment, our design leads to the following counts of irrelevant cells of information displayed to subjects:

Treatment	Irrelevant Cells
O_5A_5	0
O_5A_{15}	50
$O_{15}A_5$	50
$O_{15}A_{15}$	200

Table 2.11: Treatments and Irrelevant Information

So since we find higher mistake rates in treatment $O_{15}A_{15}$ only, this could be the result of either a) convexity in the mistake rate function or b) the presence of an additional 150 irrelevant cells relative to treatments O_5A_{15} and $O_{15}A_5$. Using the alternative 8x8 design, we can more precisely investigate the effect of the "size" of the irrelevant information set on mistake rates. The 8x8 design leads to the following:

If mistake rates in treatment O_8A_8 are higher than in treatment O_5A_5 in this new dataset, we can thus conclude that this is the result of convexities in the mis-

Treatment	Irrelevant Cells
O_5A_5	0
O_5A_8	15
O_8A_5	15
O_8A_8	49

Table 2.12: Robustness Treatments and Irrelevant Information

take rate function and not simply the size of the set of irrelevant information. This is indeed the case, as displayed in Table C.17 in Appendix C.10.1: mistake rates are roughly 8 percentage points higher in treatment O_8A_8 relative to the baseline. Moreover, adding three unavailable options or three irrelevant attributes alone does not increase mistake rates relative to the baseline. Treatment O_8A_8 with 49 irrelevant cells displayed to subjects, has a mistake rate of 24.2%, higher than either O_5A_{15} and $O_{15}A_5$ in the main dataset. [12] Taken together, these additional analyses reveal that the central result contained in this work is indeed due to the presence of both unavailable options and irrelevant attributes, not simply due to the sheer amount of irrelevant information displayed.

2.4.2 Additional Willingness-to-Pay

As previously mentioned, we've shown that both unavailable options and irrelevant attributes are necessary to generate an increase in the mistake rate, with mistake rates in treatment $O_{15}A_{15}$ being significantly higher than in the baseline. We've also shown that WTP to eliminate irrelevant information is sensitive to individual mistake rates, even though subjects are not provided with feedback regarding

[12]We also view mistake rates in the treatments used for robustness as lower bounds on true mistake rates. The mistake rate for the baseline treatment of this dataset was 16.8%, lower than the baseline mistake rate of 21.3% for the main dataet.

their performance in Part 1 of the experiment prior to submitting their WTP.

To bridge these two results, we conducted an additional two sessions where Part 2 of the experiment was altered to only as a single WTP question, with subjects submitting their WTP to move from $O_{15}A_{15}$ to O_5A_5. We asked only one WTP question in these sessions to avoid priming the subjects to rank WTP in a particular order. Our central hypothesis is that, because mistake rates are higher only in $O_{15}A_{15}$, WTP for $O_{15}A_{15}$ to O_5A_5 should be significantly higher than any other WTP measure. If we had asked, say, three WTP measures ($O_{15}A_{15} \rightarrow O_5A_{15}$, $O_{15}A_{15} \rightarrow O_{15}A_5$, and $O_{15}A_{15} \rightarrow O_5A_5$) in these sessions, the subject may be primed to internally rank these three WTPs with $O_{15}A_{15} \rightarrow O_5A_5$ as the "most valuable" simply due to the relatively large number of irrelevant cells eliminated. To avoid this priming, we ask for WTP for $O_{15}A_{15} \rightarrow O_5A_5$ alone.

We find results consistent with our hypothesis, as indicated in Tables C.23 and C.24 in Appendix C.10.2. Mean WTP for $O_{15}A_{15} \rightarrow O_5A_5$ is 5.452 ECU, higher than any other WTP measure previously elicited in the main dataset. Moreover, approximately 84% of subjects submitted a positive WTP for $O_{15}A_{15} \rightarrow O_5A_5$, again higher than any other frequency elicited for the main dataset. These results provide more credence to the notion that WTP to eliminate irrelevant information closely tracks performance in Part 1, even absent any feedback.

2.5 Discussion

From the above analysis we've shown that irrelevant information can increase the frequency of sub-optimal choice. This has implications for how we model both rational choice under constraints on attention and boundedly rational choice. We can reject purely random choice in each treatment: note that mistake rates in each treatment would be equal to 80% (since one of the five available options will always be optimal) if subjects choose randomly, giving each option an equal chance of being chosen. We can reject a null hypothesis that mistake rates are equal to 80% in each treatment ($p < 0.000$ in each). Likewise, we can reject fully rational choice (under no attention constraints) at the $\alpha = 0.001$ level.

Given that our results are consistent with neither random choice nor fully rational choice, it remains to be seen whether a behavioral model that allows for sub-optimal choice is consistent with our data. As mentioned in Section 2.1, models that allow for sub-optimal choice focus on available options and relevant attributes. In limited consideration based models of choice (see e.g. Masatlioglu et al. (2012), Manzini and Mariotti (2007; 2012; 2014), and Lleras et al. (2017)), the decision-maker first creates a "consideration set" from the set of available options. If the optimal option in the set of available options does not make it into the consideration set, it will not be chosen and choice will be sub-optimal. Similarly in models of satisficing and search (e.g. Caplin et al. (2011)), the decision-maker searches through the list of available options, leaving the potential to fail to consider the optimal option displayed. In models of rational inattention (see e.g. Caplin and Dean

(2015); Matejka and McKay (2014); Sims (2003; 2006)), the decision-maker acquires information at some cost through a rational attention allocation process. In such a framework, the agent would optimally pay no attention to irrelevant information (i.e. unavailable options or irrelevant attributes). Similarly, the salience-based model of Bordalo et al. (2012; 2013; 2016) is based on relevant attributes only. In this model, attributes of a given option are weighted based on their distance from the mean value of that attribute across all goods that are available. Trivially, irrelevant attributes in such a model would have equal (zero) salience and would thusly be ignored.

To rectify our results with the extant body of literature, one would have to make considerable alterations to these models. The cost of acquiring information in a rational inattention framework, for example, would have to be modeled as dependent on the amount of irrelevant information displayed.[13] In models of search or satisficing, one would have to assume that the decision-maker either a) has a cost-of-search parameter that depends on the presence of irrelevant information or b) searches through unavailable options mistakenly with some probability. Similarly, the salience-based model of Bordalo et al. (2012; 2013; 2016) would have to be modified to allow for the presence of irrelevant attributes.

In this spirit, we propose the concept of a "presentation set" to be incorporated in more general choice theoretic models. A decision problem in such an approach would be defined as a (S, P)-tuple, with S and P as subsets of the grand set of

[13]In the same vein, there is a small, but growing body of literature on incorporating "perceptual distance" between states of nature into models of rational inattention (see Experiment 4 in Dean and Neligh (2017b)). Our results could be viewed through this lens: it is more difficult to perceive which option is optimal in the presence of irrelevant information, even though the state-space is payoff equivalent to the decision-problem without irrelevant information.

alternatives such that $S \subseteq P$. While S is the set of available options displayed to the consumer, a (weakly) larger set P is presented to the consumer, with $s \in P \setminus S$ interpreted as unavailable options. An attribute-dependent modification of this approach is straightforward. Our results suggest that choices depend on P as well as S.

Such an approach is related to the work of Salant and Rubinstein (2008). In their model, choice is affected by a "frame" which they define as including "observable information that is irrelevant in the rational assessment of the alternatives, but nonetheless affects choice." Since a "frame" is anything other than relevant information to the decision problem that can affect choice, the "presentation set" can be interpreted as a "frame". Nevertheless, this "presentation set" may trigger the DM to use a different choice procedure.

Consider the following example: a DM always optimizes (i.e. considers all options and chooses the best one) when the presentation set is equal to the set of available goods, but uses Simon's satisficing criteria for more complicated presentation sets. Further, suppose there are three available options, x, y, and z such that $U(x) > U(y) > U(z)$ for some utility function U and that $U(z) \geq \tau$, for some satisficing level of utility τ. Thus, if the DM is optimizing, she will choose x, but the DM will choose the first available option considered if following a satisficing criteria. Assume that there are two frames/presentation sets: f_1 where there is no additional information displayed other than the available goods and f_2 where x, y, and z are displayed along with unavailable goods.

Under f_1, the DM will always choose the U-maximal option, since the DM can optimize under simple frames/presentation sets. However, under f_2, the consumer will choose the first available option that she sees. Suppose the options are always displayed in the order $z - y - x$. Then the DM's choice correspondence will be as follows:

	$\{x, y, z\}$	$\{x, y\}$	$\{x, z\}$	$\{y, z\}$
$c(A, f_1)$	$\{x\}$	$\{x\}$	$\{x\}$	$\{y\}$
$c(A, f_2)$	$\{z\}$	$\{y\}$	$\{z\}$	$\{z\}$
$C_c(A)$	$\{x, z\}$	$\{x, y\}$	$\{x, z\}$	$\{y, z\}$

Table 2.13: Example: Choice Data for Salant-Rubinstein Application

In the above, as in Salant and Rubinstein (2008), given a set of frames, F, C_c is constructed such that $C_c(A) = \{x \mid \exists f_i \in F \text{ such that } c(A, f_i) = x\}$ for $c(A, f)$ as a choice correspondence under set A and frame f. Salant and Rubinstein (2008) present a γ-axiom under which if $x \in C_c(A) \cap C_c(B)$ then $x \in C_c(A \cup B)$, which is required for a choice with frames to be consistent with the maximization of some transitive, binary relation. This property is clearly violated in the above choice data (to see this easily, let $A = \{x, y\}$ and $B = \{y, z\}$).

This type of adaptive choice procedure is consistent with our data. Forty-eight (48 out of 222) of our subjects made no mistakes in the baseline O_5A_5 type decision problems (i.e. they are "simple optimizers" according to the above adaptive choice procedure). We define a violation of satisficing procedure as a subject choosing an option placed at position i when there is a higher-valued option placed at position $j < i$ (i.e. higher up on the screen). According to this definition, 5 of these 48 simple optimizers make no mistakes through violations of satisficing. Some 16 of the remaining 43 subjects make fewer than 60% of their mistakes through violations of satisficing. Thus, there is a sizeable (though minority) contingent of our sample who can be modeled as following the adaptive procedure described in the example above, but who will violate the central γ-axiom of Salant and Rubinstein (2008). [14]

2.6 Conclusion

In this paper we have presented the results of a novel experimental design to test for both i) effects of irrelevant information presented in a decision problem on choice and ii) willingness-to-pay to get rid of irrelevant information. Our main contribution is the identification of complementarities in irrelevant information presentation: both unavailable options and irrelevant attributes are necessary to generate increased mistake rates. This central result can shed light on the extant body of literature on decision theory and limited attention. Namely, we find that no leading models of choice, either rational and constrained or boundedly rational, can explain our data unless they are significantly modified. It is our hope that these results may provide direction for upcoming theoretical research intended to model choice in the presence of irrelevant information.

Our results are applicable to a number of contexts in the realms of public policy, marketing, and choice architecture. In particular, our results indicate that choice architects should possibly err on the side of simplicity when presenting information that may or may not be pertinent to all DMs who will see it. In the United States, there is currently robust debate as to whether or not the federal government should require the food industry to label goods as having genetically engineered

[14] This example is similar to the two moods example (Salant and Rubinstein, 2008, page 1294).

(GE) ingredients or not. Indeed, George Kimbrell, Legal Director for the Center for Food Safety, an American advocacy organization, has stated that "Americans deserve nothing less than clear on-package labeling [regarding GE ingredients], the way food has always been labeled" (Center for Food Safety, 2017). In the absence of scientifically proven health concern about GE, this information will be "irrelevant" but it will distract to pay less attention to the relevant attributes, such as sodium or fat contents of the food. Our results should inspire caution on the part of policy-makers. Armed with only a rational model of consumer choice, a policy-maker may decide that "more information is always better for consumers." Our results indicate that not only may this additional information make it more difficult to choose optimally for a consumer who finds the information irrelevant, such a consumer may simply have an unexpressed preference for simplicity. We see similar applications of these results in such areas as prescription drug labelling.

Returning to the example presented in the Introduction, online shopping platforms such as Amazon.com, Wayfair.com, and Jet.com appear to be unsure of how to treat the unavailability of certain goods. For example, even within Amazon's platform, there is contrasting treatment of out-of-stock goods: for standard goods, Amazon displays no out-of-stock options by default, allowing the consumer to opt into seeing out-of-stock items, but when searching on Amazon Fresh, Amazon's grocery delivery platform, out-of-stock items are displayed by default with no option to opt out of seeing them. We should caution that our results don't suggest that not displaying such information is always optimal for the firm; displaying such information may be profitable for a number of reasons, including dynamic alternative sets and purchasing decisions, reference dependence (away from which we have abstracted in this work),such as the possibility that unavailable goods serve as decoy options that make certain available goods seem more attractive. However, our results do suggest that any agent considering whether to display such irrelevant information should recognize that there is a trade-off: a firm must weigh the potential immediate effect on profit relative to the effect on choice optimality on the part of the consumer that is induced by the presence of irrelevant information.

Further, we identify a pure “preference for simplicity”. That is, for a subject who is faced with no cognitive costs of having to ignore irrelevant information, we find that they are still willing to pay some amount to get rid of this information. This tells us that there are aspects of consumer preference in this environment that are not fully contained by measures intended to capture the notion of lost monetary value (i.e. mistake rates and time required to make a decision). It needs to be further investigated in future research how the complexity of presentation affects the algorithm used in decision making and how robust the preference for simplicity we document here is with respect to features of the decision problems used, such as color coded irrelevant information.

Chapter 3: Reciprocity and Uncertainty: A Laboratory Experiment of Gift Exchange with Risk

Co-authored with Erkut Ozbay

3.1 Introduction

Positive reciprocity - responding to a positive action taken by another agent by taking a positive action in return - is a well-documented behavior in both the lab and the field. Laboratory studies of positive reciprocity in the form of "gift exchange" have consistently identified Pareto-improving trade among pairs of subjects when such trade is not predicted under the assumption of rational self-interest. Behavior documented in the lab serves as a baseline for observed reciprocity in the field. However, while reciprocal behavior in the real world can take many forms, some including some element of uncertainty, the vast majority of laboratory studies of gift exchange are conducted in environments with no uncertainty.

Take the following scenario as an example: there is a firm who is considering how many stock options to offer to a worker as part of their compensation package. Once the worker receives this compensation package, they can then choose to work hard on the job or shirk. The returns to the worker exerting high effort are going to be subject to some uncertainty, both aggregate (e.g. "what if the stock market crashes tomorrow?") and idiosyncratic (e.g. "we've lost power to our main plant and cannot produce for a month"). Similarly, the value of the stock options offered to the employee will be subject to both aggregate and idiosyncratic risks.

All of this begs several questions. Is positive reciprocity a meaningful behavioral phenomenon in environments with uncertainty? If so, are there differences in the form that such reciprocity takes relative to environments with no uncertainty? Finally, given that essentially all of the available results in controlled laboratory environments come from environments with no uncertainty, how well can these proxy for behavior in situations with risk?

To answer these questions, we conducted a laboratory experiment using an extension of the standard Gift Exchange Game to environments with uncertainty: the Stochastic Gift Exchange Game (SGE). In SGE, a Firm and Employee are both given 100 lottery tickets and can exchange them with one another. The Firm first offers a "wage" to the employee and the Employee then responds with "effort." At the end of this game, one prize can be won by either the Firm or the Employee, but not both. In this way, we shut down the possibility of ex-post reciprocal concerns, that is, caring about giving in terms of the final allocation between the agents only, operating as a meaningful channel for Gift Exchange. In contrast, only "ex-ante" reciprocal concerns, or caring about giving in expectation, in SGE could generate positive levels of reciprocity. We find that this is the case: similar to deterministic studies of the Gift Exchange game, Wages offered are significantly higher than the subgame-perfect Nash equilibrium prediction and these higher Wages are rewarded with above-equilibrium levels of Effort. Moreover, Effort provision is sensitive to the size of the Wage, similarly in line with previous results regarding deterministic gift exchange.

We also utilize a within-subject design to elicit behavior in the Deterministic Gift Exchange game (DGE) for each of our subjects. While there are positive levels of giving in SGE, Wages are significantly lower in SGE than in DGE, indicating a systematic difference in the form of positive reciprocity in environments with risk. Because real world environments seldom are totally absent risk and because laboratory studies of reciprocity are typically conducted for DGE only, we ask whether giving in DGE is a decent enough proxy for giving in SGE. We find that it isn't: giving in DGE only explains a portion of giving in SGE. This presents a problem for the extant body of literature on reciprocity and could be indicative of an overestimation of the amount of reciprocal giving in the real world by undue reliance on deterministic environments in the lab.

This work extents the current body of literature on gift exchange to environments with uncertainty. Beginning with Fehr et al. (1993), laboratory studies of variants of the Gift Exchange game have consistently documented above equilibrium wages and effort, as well as sensitivity of effort to the size of the wage. [1] Gift exchange is robust to modifications in the sequence of strategy choice (Carpenter, 2017), the level of competition (Brandts and Charness, 2004), and the level of control in the environment , i.e. whether DGE was conducted as a field study

[1] See also Fehr et al. (1998), Falk et al. (1999), Gächter and Falk (2002), and others.

(Bellemare and Shearer, 2009; Falk, 2007; Gneezy and List, 2006).[2] Yet to be explored, however, is the presence of positive reciprocity in environments with risk. Our work is the first study to address this question explicitly.

This isn't to say that no variants of DGE exist with some element of stochasticity. Charness and Levine (2007) introduces stochasticity in wages, not to test the effect of stochasticity on the presence of positive reciprocity directly, but to test whether positive intentions matter distinctly from the ex-post outcome. A related design is used in both Rubin and Sheremeta (2015) and Davis et al. (2017). They find that employees, for example, reward medium wages more when a higher wage was chosen by the firm, but "bad luck" in the form of a negative shock lowered the actual wage received by the employee than when the medium wage itself was chosen by the firm. However, in their experiment, the choice of a wage always affected both the ex-ante payoff distribution and ex-post payoff distribution for the two agents. In contrast, in SGE, a choice of a positive wage can only affect the ex-ante allocation between the two agents, always in the direction of increasing the probability of an ex-post unfair allocation. To our knowledge, ours is the first study to explicitly consider the presence of an ex-ante reciprocal motive.

Our work herein is closely related to the growing body of literature at the intersection of social preferences and risk preferences, particularly studies of ex-ante fairness motives. The experiment most similar to ours is that of Brock et al. (2013). They present subjects with a stochastic version of the Dictator Game where a dic-

[2]One exception is Charness et al. (2004), where they found that the absence of payoff tables presented in experimental instructions effectively eliminated positive reciprocity.

tator gives a certain number of tokens to a recipient, with each token representing a chance at some monetary prize. They find that theories of fairness concerns that do not simultaneously account for ex-ante and ex-post motives are incapable of explaining their data. Fudenberg and Levine (2012) were the first to identify the need for a theory of ex-ante fairness, with Saito (2013) offering a tractable Expected Inequality Aversion (EIA) model shortly thereafter. EIA extends the work of Fehr et al. (1993) by proposing a value function that is the weighted sum of the Fehr et al. (1993) Inequality Aversion utility function applied to ex-ante and ex-post concerns separately. A related approach is taken in López-Vargas (2015), where the value function is discounted by a function of the distance between two agents' distribution of payoffs. These two approaches, though distinct, share the feature that ex-ante and ex-post concerns are separable and fundamentally operate according to the same psychological channel. Our results are not in accordance with this interpretation of ex-ante preferences and this is discussed in Section 3.4.3. In particular, we find evidence of an endowment effect on the ability to affect either the ex-ante or ex-post outcome, suggesting that incorporating some form of reference dependence a la Tversky and Kahneman (1991) may provide a fruitful approach to modeling ex-ante reciprocity.

The rest of the paper is organized as follows. In Section 3.2 we present the experimental design and findings regarding the presence of ex-ante reciprocity. We discuss the implications of our results for both the Gift Exchange body of literature and the larger body of literature at the intersection of social and risk preferences in Section 3.4. Section 3.5 concludes.

3.2 Experiment

We ran 16 sessions of the experiment at the Experimental Economics Laboratory at the University of Maryland, College Park. In each session, 16 subjects participated, for a total of 256 subjects. Subjects earned an average of $22.25 USD. Due to the stochastic nature of payoffs, a function of our experimental design, there was considerable heterogeneity in cash earnings: the minimum payoff was $8 USD; the maximum payoff was $45 USD; and the standard deviation was $8.66. Sessions lasted approximately 90 minutes. The experiment was programmed in zTree Fischbacher (2007).

In each session, subjects completed 5 tasks: i) a Stochastic Gift Exchange Game which varied by treatment, ii) the Deterministic Gift Exchange Game, iii) Self Holt-Laury (2002) risk elicitation procedure, iv) Other Holt-Laury (2002) risk elicitation procedure, and v) Questionnaire(s). The order was fixed, moving from Task One to Task Five in each session. Tasks are described in detail below.

3.2.1 Stochastic Gift Exchange

In the first task, subjects played eight one-shot Stochastic Gift Exchange Games (SGE). For each SGE, a subject was randomly matched with another subject in the lab and the SGE was played in pairs. No subject played the SGE with the same partner more than once. Subjects in the first eight sessions played the standard SGE described in this section. In the remaining eight sessions, subjects completed one of two variations on the standard SGE, explained in Section 3.3.3.

The Stochastic Gift Exchange game was explained to subjects as follows:

There are two players, a Firm and an Employee. [3] Roles were assigned to subjects in the first period and remained fixed for each of the eight SGE that were played ina given session. The game contains two stages: Stage One in which the Firm makes a decision and Stage Two in which the Employee makes a decision. In Stage One, the Firm starts with 100 virtual "tokens" and can choose how many they would like to give to the Employee. The Employee then receives five times that number of tokens. In Stage Two, the Employee is given 100 additional tokens and can choose how many (out of 100) they would like to give to the Firm. The Firm then receives five times the number of tokens given to them by the Employee.

Each token held by the Firm or Employee at the end of Stage Two represents a $\frac{1}{1000}$ chance at a single prize of \$20. The prize can be won by the Firm, the Employee, or by neither; there is no event wherein both the Firm and Employee win a monetary prize. Note that the initial endowments of 100 tokens each implies an ex-ante endowment of a 10% chance of winning the prize, with the remaining 80% held by the experimenter. At the end of Stage Two, the expected payoff for each player $i \in \{Firm, Employee\}$ is as follows:

$$\pi_i(\tau_i^t, \tau_j^t) = 20 \cdot \frac{100 - \tau_i^t + 5\tau_j^t}{1000} \tag{3.1}$$

where τ_i^t refers to the chosen transfer of tokens from player i to player j.

Note that only by transferring the entirety of both endowments from one player

[3] The terminology "Person 1" and "Person 2" were employed in our instructions for the roles of "Firm" and "Employee," respectively.

to the other may a pair ensure that the prize would be won by either the Firm or Employee. In this sense, trade is ex-ante Pareto Optimal, a feature shared by both the Deterministic Gift Exchange game described in 3.2.2 and in other studies of reciprocity that share a similar payoff structure.

In the SGE Task, subjects completed eight rounds of the SGE game for their treatment using perfect stranger matching. Each subject kept the same role (either Firm or Employee) for all eight rounds.

3.2.2 Deterministic Gift Exchange

The Deterministic Gift Exchange (GE) game was standard and works as described above for SGE, except that payoffs were deterministic. Instead of receiving 100 tokens, subjects were given 100 Experimental Currency Units (ECU). Again, in Stage One, the Firm chose how many of 100 ECU they would like to transfer to the Employee. The Employee was then given five times that number of ECU. In Stage Two, the Employee chose how many of 100 ECU they would like to transfer back to the Firm. At the end of Stage Two, the number of ECU held by player i is as follows:

$$\pi_i(\tau_i^{ecu}, \tau_j^{ecu}) = 100 - \tau_i^{ecu} + 5\tau_j^{ecu} \tag{3.2}$$

where τ_i^{ecu} refers to the chosen transfer of ECU from player i to player j. At the end of this Task, ECU was converted to USD at a rate of 50 ECU to \$1 USD. Note that this exchange rate implies that if a given pair choses strategy profiles in SGE and DGE such that $(\tau^{t}_{Firm}, \tau^{t}_{Employee}) = (\tau^{ecu}_{Firm}, \tau^{ecu}_{Employee})$, it would lead to equal expected utilities across SGE and DGE for a risk-neutral agent.

Subjects completed one round of DGE in Task Two and were randomly matched with another participant in the room. They had the same roles as were assigned previously for SGE.

3.2.3 Self Holt-Laury Risk Elicitation Task

After completing both the SGE and DGE tasks, subjects were asked to complete the risk preference elicitation task described in Holt and Laury (2002). In this task, the subject was to make ten decisions, choosing between two options in each decision. Each decision had a "safe" option and a "risky" option, as given in the original Holt and Laury (2002) experiment. At the end of this task, two lotteries (per subject) were conducted: one to decide which of the ten decisions to use for payment, and one to realize the outcome of the subject's choice for that decision. Subjects were then informed of their payoff from this task.

3.2.4 Other Holt-Laury Risk Elicitation Task

Because SGE involves strategy choices that i) involve risk and ii) involve another agent's payoffs, we need to control for potential "other-regarding" risk preferences. That is, it is possible that a subject's choices among risk alternatives may depend on whether that choice is made for oneself or <u>on behalf of</u> another person. To this end, we repeated the risk preference elicitation task of Holt and Laury (2002), but instead informed subjects that they would be choosing on behalf of another subject in the lab. They were also informed that another subject in the lab would be choosing on their behalf, but that these people (i.e. the person for whom subject X is choosing and the person choosing on behalf of subject X) may not be the same. This matching was completed anonymously and such that every subject made decisions for one other subject in the room, but without explicit partner matching.[4]

3.2.5 Questionnaire

Finally, each subject completed a short demographic questionnaire. Subjects were asked about their age, gender, self-reported SAT and ACT scores, and their GPA. In addition, there was an open-ended field where subjects could give some narrative justification for the decisions that they made in the experiment.

3.3 Results

Our subject pool is balanced in terms of age and gender. The average age of our subjects was 20.125 years and 45.31% were female. Subject statistics, including self-reported SAT, ACT, and GPA , are reported in Table 3.1 below.

As mentioned previously, there are several motivations for studying the presence of reciprocity in an environment with uncertainty. We consider each of them in turn in the following subsections.

[4]Specifically, each subject was randomly assigned an id number from 1 to 16 in the experimental program. Subject i then made decisions on behalf of subject $i+1$, with Subject 16 making decisions on behalf of Subject 1. These id numbers were never revealed to subjects during this task.

Table 3.1: Subject Statistics

	mean	sd	min	max	count
Age	20.125	1.611235	16	29	256
Female	.453125	.498773	0	1	256
SAT	1844.976	308.8474	1000	2400	210
ACT	29.91346	3.829144	18	36	104
GPA	3.364882	.4309551	1.7	4	254
Observations	256				

SAT, ACT, GPA observations were dropped if the subject submitted infeasible responsed 4, 29, and 2 observations dropped for SAT, ACT, and GPA, respectively

3.3.1 Does Ex-Ante Reciprocity Exist?

Leading models of reciprocity, including Falk and Fischbacher (2006), Bolton and Ockenfels (2000), and Cox et al. (2007), do not predict the presence of any gift exchange in our environment when incorporated into an expected utility model. More precisely, these theories can lead to positive wages and effort, but imply no sensitivity of effort to the proffered wage.

From Figure 3.1 below, we can see both positive wages offered by the Firm in SGE and positive Effort in response. There is also little evidence of learning: no discernible trend in Wage or Effort by Period can be detected. Finally, offered Wages, but not Effort, are somewhat lower on average in SGE relative to DGE. These differences are explored further in Section 3.3.2

Provided Effort levels are also sensitive to the Wage offered in SGE. Figure 3.2 shows an overall positive relationship between offered Wage and Effort in response in SGE: average Effort is roughly 10 tokens in response to Wages offered between 0 and 19 tokens, which increases to roughly 58 tokens in response to offered Wages between 80 and 100 tokens. This effect remains when controlling for subject-level

Table 3.2: Wages and Effort: SGE and DGE

	SGE	DGE
Wage		
Mean	40.420	51.250
Std Error	1.595	1.722
Min	0	0
Median	30	50
Max	100	100
N	512	64
Effort		
Mean	27.305	32.609
Std Error	1.474	1.694
Min	0	0
Median	10	10
Max	100	100
N	512	64

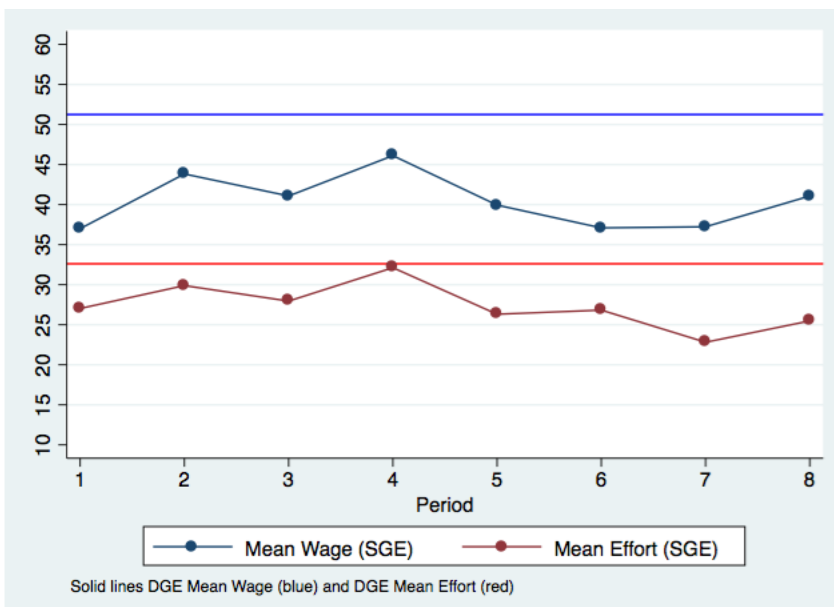


Figure 3.1: Mean Wage and Effort by Period

heterogeneity in the tobit regression specifications given in Table 3.6: the coefficient on Wage is positive and significant in each model specification where the dependent variable is Effort. Furthermore, the coefficient on Wage is significantly greater than 0.2, indicating that a strictly positive Wage was profitable in SGE in expectation.

Table 3.3: Determinants of Effort: SGE and DGE

	Model 1	Model 2
Wage	0.720***	0.711***
	(0.210)	(0.215)
Stochasticity	-3.063	-3.549
	(8.607)	(8.760)
Wage * Stochasticity	-0.0347	-0.0259
	(0.219)	(0.221)
Period	-1.329*	-1.333*
	(0.802)	(0.800)
GPA		-3.314
		(4.934)
Female		1.159
		(7.710)
Constant	3.143	13.97
	(9.040)	(20.11)
sigma		
Constant	40.75***	40.71***
	(3.823)	(3.851)
Observations	576	576

Standard errors in parentheses

Tobit regressions with lower limit equal to 0 and upper limit equal to 100

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

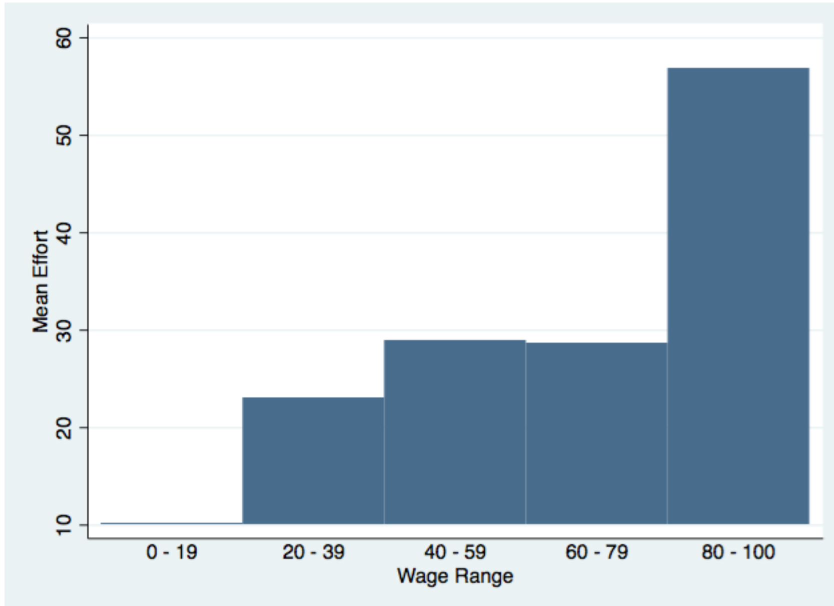


Figure 3.2: Mean Effort by Wage: SGE

Taken together, the analysis above leads us to our first result:

Result 10. *Ex-ante reciprocal motives exist in SGE:*

- *Mean Wages and Effort are both positive in SGE*
- *Effort provided by Employees is sensitive to the size of the Wage in SGE*
- *Positive Wages are profitable in expectation*

3.3.2 Does Gift Exchange Differ Between SGE and DGE?

Having established that gift exchange does indeed exist in SGE, we now turn to explicit differences between Wages and Effort between SGE and DGE. There are several reasons why we might expect differences in either Wages or Effort between the two environments. First, SGE involves an element of uncertainty not present in DGE: one token transferred may not actually result in a change in the ex-post allocation of the recipient. As such, we might expect risk preferences to have an effect on Wages and Effort in SGE, but not in DGE. Additionally, subjects are only able to choose transfers based on ex-ante reciprocal motives in SGE, since no transfer can chance the set of feasible ex-post allocations of the two players. In DGE, ex-ante and ex-post reciprocal motives coincide: a transfer of one ECU from one player to the other increases the allocation of the recipient in expectation (trivially) and in actuality. Thus, if ex-post reciprocity matters, we may expect lower levels of Wages or Effort in SGE relative to DGE.

Mean Wages and Effort are given in Table 3.4. On average, subjects offer about 10.83 fewer tokens in the Wage in SGE relative to DGE (significant at the $\alpha = 0.05$ level). However, overall Effort levels are not different between the two environments. These effects survive in tobit regression specifications in Tables 3.5 and 3.6, which control for potential learning (i.e. controlling for Period) and individual heterogeneity. As seen in Table 3.5, subjects with higher GPAs offer higher Wages and Female subjects offer fewer, on average.[5] The same is not true for Effort provided in response. Additionally, there appears to be no learning in terms of sensitivity of Wage to Period; however, Effort provided does decrease with each Period.

This fundamental difference in Wages offered between SGE and DGE is not limited to simply the mean. As can be seen in Table 3.7, more subjects offered a Wage of zero in SGE, though this difference is not statistically significant. Addition-

[5]We view the detection of lower Wages offered on the part of our Female subjects as merely the result of controlling for gender in our regression specifications, not as a fundamental result about gender differences in giving. Ex-ante fairness concerns, ambiguity attitudes, and other preferences may differ between men and women, none of which is controlled for in these specifications.

Table 3.4: Mean Wage and Effort by Stochasticity

	DGE	SGE	Difference
Wage	51.25	40.42	-10.83**
	(4.872)	(1.595)	
	64	512	
Effort	32.61	27.30	-5.305
	(4.168)	(1.474)	
	64	512	

Std. Errors in Parentheses

Table 3.5: Determinants of Wage: SGE and DGE

	Model 1	Model 2
SGE	-21.98***	-21.55***
	(7.635)	(7.577)
Period	-0.843	-0.804
	(0.872)	(0.856)
GPA		37.83***
		(11.54)
Female		-19.08**
		(8.951)
Constant	67.11***	-51.32
	(11.68)	(37.42)
sigma		
Constant	48.69***	45.05***
	(4.587)	(3.812)
Observations	576	576

Standard errors in parentheses

Tobit regressions with lower limit equal to 0 and upper limit equal to 100

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.6: Determinants of Effort: SGE and DGE

	Model 1	Model 2
Wage	0.720***	0.711***
	(0.210)	(0.215)
Stochasticity	-3.063	-3.549
	(8.607)	(8.760)
Wage * Stochasticity	-0.0347	-0.0259
	(0.219)	(0.221)
Period	-1.329*	-1.333*
	(0.802)	(0.800)
GPA		-3.314
		(4.934)
Female		1.159
		(7.710)
Constant	3.143	13.97
	(9.040)	(20.11)
sigma		
Constant	40.75***	40.71***
	(3.823)	(3.851)
Observations	576	576

Standard errors in parentheses

Tobit regressions with lower limit equal to 0 and upper limit equal to 100

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

ally, fewer subjects offered half of their endowment or all of their endowment (i.e. Wage $= 50, 100$, respectively) in SGE. Figure 3.3 presents the CDF of Wages in SGE and DGE, indicating a difference between these two distributions in the direction of lower Wages overall in SGE. While Table 3.7 seems to indicate that fewer subjects responded with Effort of 50 or 100 in SGE, Figure 3.4 shows that these two CDFs are not statistically difference from one another.

Table 3.7: Wage and Effort Frequencies: SGE and DGE

	DGE	SGE	Difference
Wage			
0	0.0469	0.109	0.0625
	(0.0266)	(0.0138)	
	64	512	
50	0.188	0.107	-0.0801*
	(0.0387)	(0.0137)	
	64	512	
100	0.344	0.164	-0.180***
	(0.0598)	(0.0164)	
	64	512	
Effort			
0	0.312	0.250	-0.0625
	(0.0542)	(0.0192)	
	64	512	
50	0.156	0.0840	-0.0723*
	(0.0457)	(0.0123)	
	64	512	
100	0.188	0.0957	-0.0918**
	(0.0368)	(0.0130)	
	64	512	

Std. Errors in Parentheses

Finally, given that there are meaningful differences between Wages offered in environments with uncertainty (SGE) relative to those with none (DGE), it may be a concern that the bulk of laboratory studies of gift exchange exclusively focus on variants of DGE. Moreover, given that compensation schemes in the real world

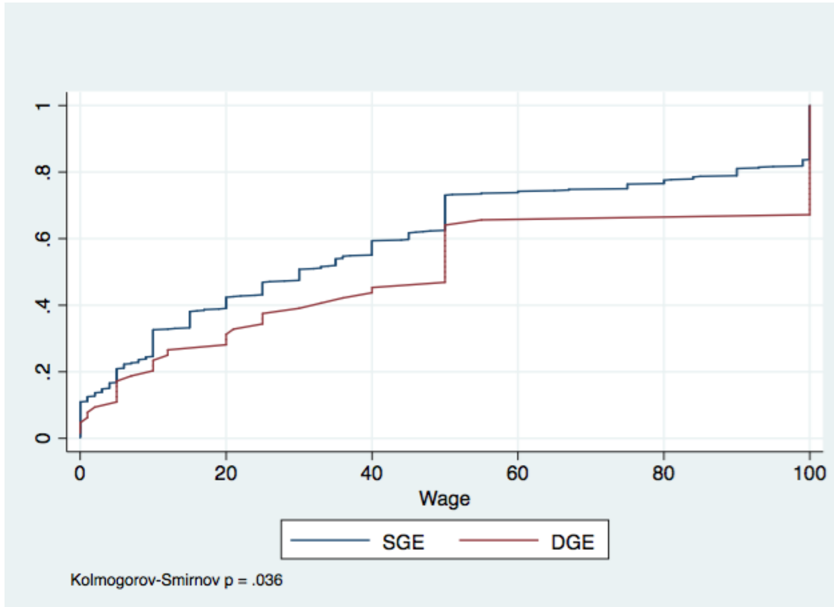


Figure 3.3: Wage CDF: SGE and DGE

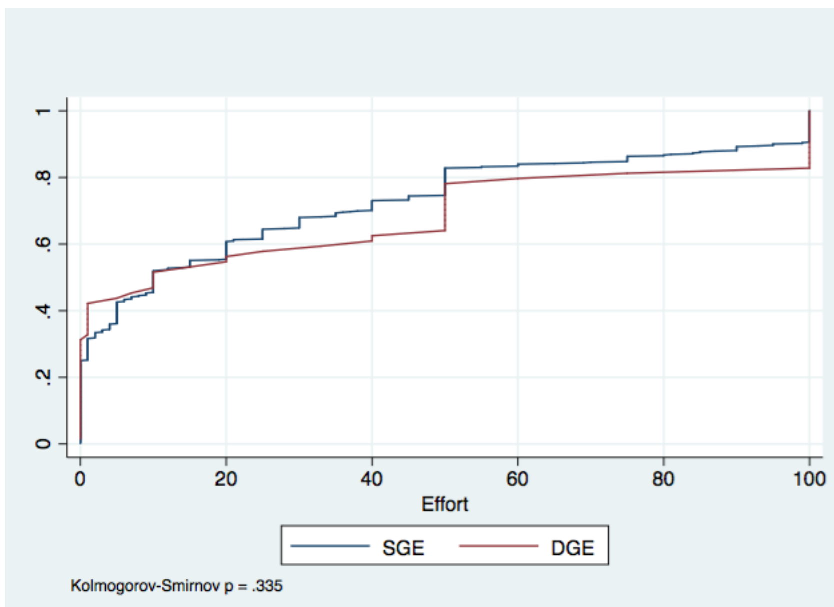


Figure 3.4: Effort CDF: SGE and DGE

can involve stochastic elements (e.g. stock options, equity, retirement benefits) and that the returns to effort provided by some worker will always be subject to both idiosyncratic and aggregate shocks, this points to a weakness of the current body of literature on gift exchange in terms of external validity. we then ask to what extent measures of reciprocity taken in DGE can proxy for behavior in SGE. The results in Table 3.8 are not promising in this regard. These tobit regressions were conducted with Wage in SGE as the dependent variable, using Wage in DGE by the same subject as the primary independent variable. We control for both learning (i.e. Period) and individual heterogeneity by GPA and gender. In neither model is the coefficient on Wage DGE close to 1, which would indicate that Wages in DGE are a perfect proxy for Wages in SGE. The coefficient is positive and significant, but much variation in Wages under SGE is left unexplained by Wages in DGE alone. [6] The degree to which Wages in SGE vary even controlling for Wages in DGE can further be seen in Figure 3.5. Confidence intervals at on the 95% level are quite wide and do not indicate much precision in the estimation of Wages in SGE using Wages in DGE alone.

Taken together, the analysis in this section leads us to our next three results:

Result 11. *Wages are lower in SGE than in DGE:*

- *Mean Wages in SGE are lower than in DGE*
- *Cumulative distribution functions are different between SGE and DGE in the direction of lower Wages in SGE*
- *There are fewer observations of Wages at* 50 *and* 100 *in SGE than in DGE*

Result 12. *Effort levels are equal across SGE and DGE.*

[6] In an ideal setting, we could report the R^2 from these regression specifications as a measure of "variance in Wages under SGE explained by a DGE proxy." However, tobit regressions, which are required for this analysis due to the upper and lower limits on the strategy space for our subjects, have no R^2 statistic. Pseudo-R^2s can be reported, but offer no similar interpretation in this setting.

Table 3.8: DGE as Proxy for SGE

	Model 1	Model 2
DGE Give	0.682***	0.560***
	(0.108)	(0.128)
Period	1.029	-1.491*
	(0.858)	(0.904)
GPA		6.084**
		(2.754)
Female		0.646
		(7.514)
sigma		
Constant	41.01***	39.74***
	(4.165)	(3.809)
Observations	512	512

Standard errors in parentheses

Tobit regressions with lower limit equal to 0 and upper limit equal to 100

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

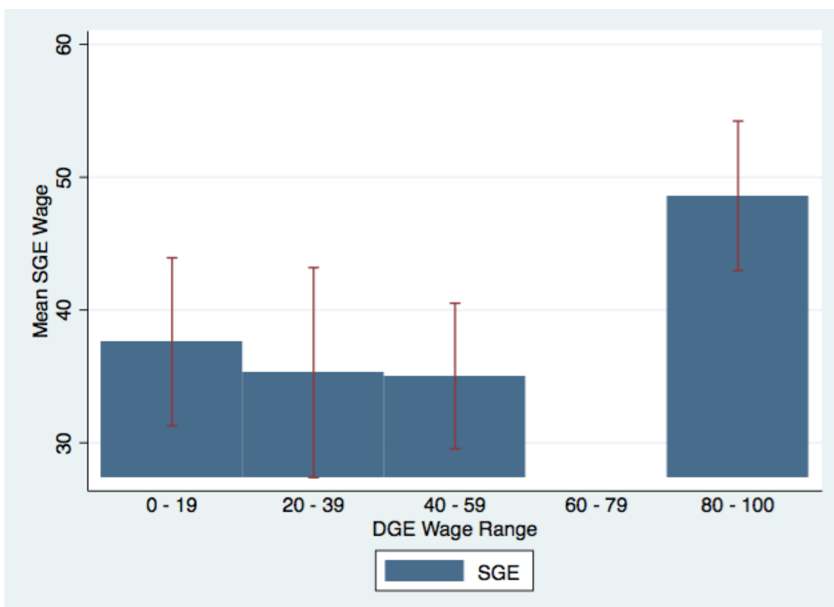


Figure 3.5: Mean Wages in SGE by DGE Wage Range

Result 13. *Wages in DGE are a poor proxy for Wages in SGE:*

- *The coefficient for "DGE Wage" with "SGE Wage" as the dependent variable is significantly less than one in all relevant regression specifications*

- *Conditional on Wages in DGE, there is significant variation in Wages in SGE for the same subject*

3.3.3 Does the Source of Risk Matter?

Given that i) reciprocity exists in the presence of risk and ii) there is a systematic difference in reciprocal behavior between environments with risk (SGE) and environments with none (DGE), it is natural to ask whether the <u>source</u> of uncertainty matters. In SGE, both Wages and Effort are subject to uncertainty: giving one token only increases the <u>expected</u> payoff of the recipient, but cannot affect it for sure. Moreover, there is a fundamental asymmetry in SGE due to its sequential nature (i.e. Wages are offered <u>before</u> an Effort response is chosen).

In order to investigate whether the source of risk matters, we implemented four sessions each of two variants on the SGE: a Wage-SGE and an Effort-SGE. Recall that in SGE, both the Firm and Employee were endowed with 100 tokens, but in DGE both were endowed with ECU. Both players were endowed with the same good. In the two variants of SGE that we conducted, we instead endowed the Firm and Employee with <u>different</u> goods with one receiving tokens and the other receiving ECU. ECU were converted to cash directly at a rate of $1 USD to 50 ECU, just as in DGE. Tokens each gave a chance at a $20 prize that could be won exclusivelyby either player, just as in SGE. Table 3.9 below describes these endowments and the expected payoffs of each player conditional on transfers τ_j^i, denominated in USD.

These asymmetric endowments contribute to asymmetric abilities to affect the

Treatment	Endowment		Expected Payoffs	
	Firm	Employee	Firm	Employee
Wage-SGE	100 tokens	100 ECU	$20 \cdot \frac{100-\tau^{t}_{Firm}}{1000} + \frac{5\tau^{ecu}_{Employee}}{50}$	$20 \cdot \frac{5\tau^{t}_{Firm}}{1000} + \frac{100-\tau^{ecu}_{Employee}}{50}$
Effort-SGE	100 ECU	100 tokens	$20 \cdot \frac{5\tau^{t}_{Employee}}{1000} + \frac{100-\tau^{ecu}_{Firm}}{50}$	$20 \cdot \frac{100-\tau^{t}_{Employee}}{1000} + \frac{5\tau^{ecu}_{Firm}}{50}$

Table 3.9: Wage- and Effort-SGE Endowments and Payoffs

ex-ante and ex-post allocations between the two players. In Wage-SGE, the Firm could affect only the ex-ante allocation of the two players, whereas the Employee could affect both the ex-post allocation and, trivially, the ex-ante allocation. The reverse is true in Effort-SGE. Does this asymmetry lead to different Wage or Effort levels in these variants relative to SGE? We answer this and related questions in this section.

Table 3.10 below displays Wage and Effort levels for all four variants of gift exchange that we utilize in this study. Overall, Wage and Effort are lower in variants of SGE than in DGE. This effect largely remains when we control for individual-level heterogeneity, risk preferences, learning, and the size of the Wage offered in Table 3.11. In Table 3.11, Safe HL refers to the number of "safe" options chosen in the Holt and Laury (2002) Task, while Safe Other HL is the analogous measure for our variant where the subject chooses on behalf of another person. Neither Wages nor Effort are affect by this measure of risk preferences. Notably, Effort is no lower in SGE than in DGE once controlling for the size of the Wage (which is lower in SGE than in DGE, as previously mentioned).

Table 3.10: Wage and Effort: All Treatments

Full SGE	
Wage	40.42
	(1.595)
	512
Effort	27.30
	(1.474)
	512
Wage SGE	
Wage	42.59
	(2.386)
	256
Effort	18.84
	(1.710)
	256
Effort SGE	
Wage	41.68
	(2.403)
	256
Effort	23.80
	(2.063)
	256
DGE	
Wage	54.77
	(3.368)
	128
Effort	34.16
	(3.398)
	128

Std. Errors in Parentheses

Table 3.11: Treatment Effects

	No Risk Preferences		Risk Preferences	
	Wage	Effort	Wage	Effort
Full SGE	-22.10***	-4.352	-21.83***	-7.266
	(6.362)	(6.070)	(7.262)	(7.167)
Wage SGE	-26.23***	-23.55***	-27.89***	-32.16***
	(9.816)	(7.269)	(10.25)	(8.587)
Effort SGE	-21.77**	-13.06*	-19.60*	-16.83**
	(9.379)	(7.618)	(10.25)	(7.936)
DGE	0	0	0	0
	(.)	(.)	(.)	(.)
GPA	18.49**	-4.234	16.61*	-3.228
	(8.288)	(4.555)	(8.865)	(5.350)
Female	-24.78***	5.884	-30.40***	5.146
	(7.335)	(6.245)	(8.071)	(7.253)
Period	-0.201	-1.555***	0.150	-2.146***
	(0.600)	(0.570)	(0.660)	(0.649)
Wage		0.640***		0.690***
		(0.0605)		(0.0665)
Safe HL			0.539	-0.753
			(2.859)	(2.292)
Safe Other HL			-0.318	0.526
			(2.289)	(1.927)
Constant	12.94	18.09	18.48	20.06
	(27.92)	(16.39)	(32.07)	(18.40)
sigma				
Constant	50.91***	43.31***	52.18***	44.80***
	(3.165)	(3.371)	(3.563)	(3.899)
Observations	1152	1152	999	990

Standard errors in parentheses

Tobit regressions with lower limit equal to 0 and upper limit equal to 100

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

From Table 3.10, we can see that Wage levels are roughly the same across all treatments with some uncertainty, ranging from 41.68 on average in Effort-SGE to 42.59 on average in Wage-SGE, with SGE in the middle at 40.42. None of this variation is statistically significant (Mann-Whitney $p > 0.10$ in each case). There is, however, significant variation in mean Effort across treatments with uncertainty. Effort is lower on average in both Wage-SGE and Effort-SGE than in SGE (Mann-Whitney $p < 0.01$ for Wage-SGE vs SGE; Mann-Whitney $p < 0.05$ for Effort-SGE vs SGE). However, effort is not significantly different between Wage-SGE and Effort-SGE (Mann-Whitney $p > 0.10$). These differences in Effort levels are also present in the entirety of the Effort distribution. Figures 3.6 and 3.7 display CDFs for Effort between SGE and Wage-SGE and Effort-SGE, respectively. Both are significantly different in the direction of lower overall effort in both Effort-SGE and Wage-SGE. However, when controlling for the size of the Wage, Effort levels in Effort-SGE are not significantly different than in SGE, as displayed in Table 3.12 below, though the coefficient is still negative. Taking these results together, we say that Effort is weakly lower in Effort-SGE than in SGE. This difference is strict for Wage-SGE relative to SGE.

Taken together, the analysis above gives us our next results:

Result 14. *The source of uncertainty does not matter for Wages and Effort:*

- *Wages are lower in SGE, Wage-SGE, and Effort-SGE relative to DGE, equally and regardless of the source of uncertainty*

- *Effort levels are equally lower in both Wage-SGE and Effort-SGE relative to*

Table 3.12: Effort Across SGEs

	No Risk Preferences	Risk Preferences
Wage	0.608***	0.644***
	(0.0591)	(0.0661)
Full SGE	0	0
	(.)	(.)
Wage SGE	-18.92***	-24.75***
	(6.682)	(8.114)
Effort SGE	-8.412	-9.445
	(8.174)	(8.854)
GPA	-5.283	-4.273
	(4.714)	(5.581)
Female	7.174	6.806
	(6.134)	(7.152)
Period	-1.500***	-2.062***
	(0.555)	(0.628)
Safe HL		-0.790
		(2.317)
Safe Other HL		0.299
		(1.892)
Constant	17.95	18.94
	(15.28)	(18.08)
sigma		
Constant	41.24***	42.83***
	(3.366)	(3.906)
Observations	1024	880

Standard errors in parentheses

Tobit regressions with lower limit equal to 0 and upper limit equal to 100

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

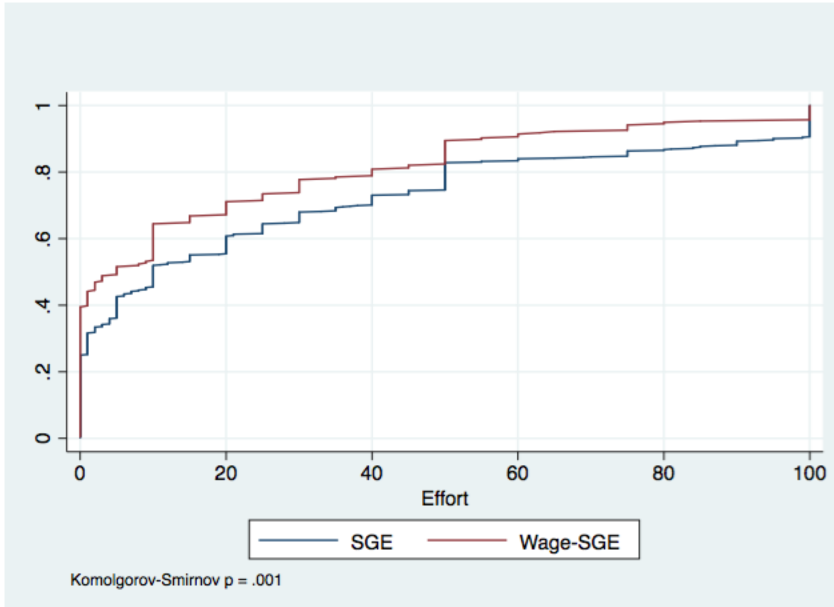


Figure 3.6: Effort CDF: Wage-SGE and SGE

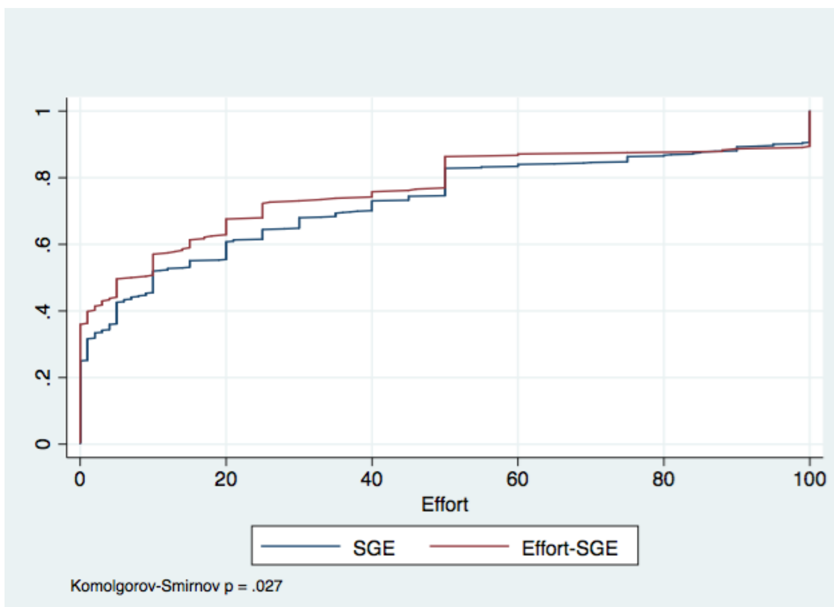


Figure 3.7: Effort CDF: Effort-SGE and SGE

DGE

Result 15. *Asymmetries in uncertainty matter for reciprocal Effort:*

- *Effort levels are lower in Wage-SGE and (weakly) Effort-SGE relative to SGE*

3.4 Discussion

In the above we have shown that i) gift exchange exists in environments with uncertainty; ii) there are significant differences in Wages and Effort offered in environments with uncertainty relative to those with none; and iii) the source of uncertainty does not affect positive reciprocity, but the asymmetry of uncertainty in the form of endowments does. Our results have implications for both how scholars should interpret laboratory findings regarding positive reciprocity and how these results should feed into the development of new theories of social preferences under uncertainty. We turn to each of these implications in turn in the following subsections.

3.4.1 On Gift Exchange With Uncertainty

To our knowledge, no leading model of reciprocity, including Falk and Fischbacher (2006), Bolton and Ockenfels (2000), and Cox et al. (2007), predicts positive levels of gift exchange in SGE. This is partly due to the reliance of these models on reciprocity coming from two channels: reciprocal intentions and outcome. We interpret "reciprocity regarding outcome" as ex-post reciprocity. Due to the asymmetrical nature of the ex-post allocation and the nature of transfers in SGE, we have effectively shut down this channel for reciprocal motives: subjects have no way to either offer wages or effort to affect the ex-post outcome. Another way of stating this is that trade from one player to another can only increase the probability of an ex-post unfair outcome (i.e. one player winning $20 and the other player winning nothing). As such, models of reciprocity based exclusively on the ex-post allocation will predict no trade in our environment.

Reliance on the ex-post allocation alone is a modelling issue that had previously been identified in the gift exchange body of literature. As such, Falk and Fischbacher (2006) and related models include measures of the "intentionality" of a gift, such that a recipient can identity and reward "good intentions" that do not necessarily positively affect the ex-post allocation. We have also shut down this channel as a meaningful source of variation between SGE and DGE. A gift of a positive number of tokens in SGE can be nothing other than fully intentional, exclusively meant to increase the expected payoff of the recipient. One might expect then that Falk and Fischbacher (2006) or related models may predict positive gift exchange in SGE. However, the psychological game presented in Falk and Fischbacher (2006) is not easily modified to include uncertainty regarding the allocation, since it already includes an ad-hoc notion of uncertainty regarding the intentions of the players (through the use of higher-order beliefs). An expected utility treatment of this model to generate predictions for SGE does not generate our results.

As such, similar to the lesson of Brock et al. (2013), we view a contribution of this work as identifying a need for further generalizations of extant theories of reciprocity to include a notion of "ex-ante" reciprocal concerns. A treatment of such concerns as merely "positive intentions" as would be required in Falk and Fischbacher (2006) is not sufficient to explain both i) positive Wages in SGE and ii) sensitivity of Effort to Wages in SGE. One may think that a reasonable first pass would be to take an approach similar to that taken in Saito (2013) for social preferences under risk. However, as we propose in Section 3.4.3, such an approach would still fall short of fitting our data.

3.4.2 On The Limits of Laboratory Gift Exchange

Our results indicate that the body of literature on gift exchange in the laboratory may be significantly overestimating the amount of positive reciprocity in the real world. In the DGE, our subjects offered Wages of 54.77 ECU or roughly half of their endowment on average across all treatments. Employee subjects responded with average Effort of 34.16 ECU, roughly one third of their endowment. These Wage and Effort levels are roughly in-line with other laboratory studies of gift exchange that use the same payoff functions. Wages were 59.0 ECU on average in the "No Minimum Wage" treatment of Owens and Kagel (2010) and 50.7 ECU overall in Filiz-Ozbay et al. (2016), both of which used idential payoff functions to that used herein.[7] Effort levels are roughly 29.1 in Owens and Kagel (2010) and 25.2 in Filiz-Ozbay et al. (2016).

However, we find considerably lower Wages in treatments with uncertainty and lower Effort levels in treatments with asymmetric endowments. In the real world, compensation packages can take many forms, some including some degree of uncertainty (e.g. stock options, equity payouts, insurance benefits, etc.). Additionally, the return associated with a given Effort level will be subject to both idiosyncratic and aggregate shocks. The uncertainty associated with these two aspects of gift

[7] Brandts and Charness (2004) also use several linear payoff functions similar to ours, but Wages offered were not "dictated": Employee subjects viewed posted Wages and choose to accept/reject them in a market setting. As such, their Wage/Effort levels were elicited in a different institutional setting and are not comparable to ours.

exchange may also be asymmetric. As such, we should expect to find that gift exchange in the real world is more in-line with our results derived from treatments with uncertainty and by focusing on environments with no uncertainty, the body of literature on gift exchange in the lab may be overestimating the amount of positive reciprocity one can expect in a reasonable real-world environment.

3.4.3 On Ex-Ante vs Ex-Post Reciprocity

We have seen in the Wage-SGE and Effort-SGE treatments that Employee subjects (weakly) prefer to keep more of their endowment when they are endowed with ECU, conditional on the size of the Wage offered. For example, in Wage-SGE when a subject is given tokens but can only respond with ECU, they give less than when they can respond with tokens in kind. The analogous is true for Effort-SGE. What we've identified is a form of the endowment effect wherein subjects prefer to keep what they already own. However, notice that tokens and ECU are effectively the same good: both tokens and ECU represent lotteries over monetary prizes, with ECU merely being degenerate versions of tokens with the same expected value.

What is different between a token and an ECU is that the former can affect ex-ante allocations only, while the latter can affect both the expected payment and the ex-post allocation. In our view, we are thus documenting an endowment effect with regards to ex-ante vs ex-post reciprocity motives rather than strict goods. Consider a gift of 1 token from Firm to Employee in Wage-SGE. The Firm has increased the expected payoff of the Employee, but has not affected the set of feasible ex-post allocations. When the Employee can only respond by using ECU, they are forced to compensate ex-ante giving with ex-post giving. The reverse is true in Effort-SGE. Because we see that Effort levels in both Wage-SGE and Effort-SGE are lower than SGE, but not Wages, we argue that Employee subjects are biased toward their own endowments. Such behavior is not consistent with current interpretations of the interplay between ex-ante and ex-post fairness.

Consider one of the most widely-used models of ex-ante vs ex-post fairness, the Expected Inequality Aversion (EIA) model of Saito (2013). In this model, decision makers maximize a value function that is a weighted sum of ex-ante fairness motives and ex-post fairness motives. This value function is given by the following:

$$V(p) = \delta U[E_p(x)] + (1-\delta)E_p[U(x)] \tag{3.3}$$

where $E_p(\cdot)$ is the expected value according to some lottery p, $U(\cdot)$ is an inequality aversion utility function of Fehr et al. (1993), and $\delta \in [0,1]$ is a measure of the strength of the ex-ante fairness motive of the decision maker. Ex-ante motives are described by a preference for equality in expectation, given by $U[E_p(x)]$ above. Ex-post motives are described by the standard inequality averse utility function extended to the expected utility case, given by $E_p[U(x)]$. Any model that combines ex-ante and ex-post motives in this way will imply that actions taken taken to affect the ex-ante portion of the decision maker's value function can be directly compensated with commensurate actions to affect the ex-post portion: for any change Δ in $U[E_p(x)]$, a change of $\Delta' = \frac{\delta}{(1-\delta)}\Delta$ in $E_p[U(x)]$ will have the same effect on $V(p)$.

This implication is not borne out in our data. While Wages remain constant across SGE, Wage-SGE, and Effort-SGE, Effort levels are systematically lower in the latter two. This cannot be explained by any model that relies on ex-ante and ex-post motives being separable. Lower Effort in Wage-SGE relative to SGE can only mean that Employee subjects believe that fewer ECU are required to compensate for a given gift of tokens (i.e. Wages in SGE and Wage-SGE). This would imply a particular rate of exchange between ex-ante and ex-post motives. However, since Wages are also equal across SGE and Effort-SGE, this rate of exchange should then imply higher Effort in Effort-SGE, since the Wage now affects ex-post motives in the form of ECU. This is not the case in our data, where we find weakly lower Effort levels in Effort-SGE relative to SGE. The result is a form of the endowment effect, not on the good that makes up the endowment, but on which motive, ex-ante or ex-post, the endowment is capable of affecting. When endowed with the ability to affect only the ex-ante allocation, as in SGE and Effort-SGE, Employee subjects seem to care more about giving in expectation. When endowed with the ability to affect the ex-post allocation, as in Wage-SGE, they seem to care more about giving in terms of the final allocation.

While extant models of reciprocity and fairness incorporating ex-ante concerns cannot immediately explain our data, we view the identification of this endowment effect as being a worthwhile place to begin theoretical examinations of ex-ante reciprocity. It is possible that a model of loss aversion a la Tversky and Kahneman (1991), modified to consider “ex-ante” vs “ex-post” domains instead of the traditional “gain/loss” framework could explain our results. Similarly, models incorporating status quo bias, such as Masatlioglu et al. (2005) or Dean et al. (2017), could be extended into the an ex-ante/ex-post framework. These theoretical questions are beyond the scope of the current work, but we view them as fodder for interesting further exploration.

3.5 Conclusion

In this work, we study i) the presence of positive reciprocity in environments with risk and ii) the difference in the form of reciprocity when uncertainty is present. Our results tell us about the limits of exclusively studying gift exchange in deterministic environments in the lab, but also provide plausible routes upon which to theoretically study ex-ante reciprocity. Considering the presence of an endowment effect on reciprocal motives in our results, we view reference-dependent models as providing a decent starting point for future theoretical exploration of these questions.

Considering the differential effects of uncertainty in the transfer between Firm and Employee on expected profit and reciprocal motives, it remains to be see just what mixture of uncertainty is optimal. Should Firms lower the proportion of a compensation scheme dedicated to stock options and/or equity, given that Employees respond with lower effort in response? Should Firms invest in lowering the uncertainty in how employee effort feeds into firm profit? We view studying these and related questions as being fruitful follow-up studies to that contained herein.

A.6 Proofs For Chapter 1

Lemma 1. *For any g such that $g[S]$ is T-Connected for some $T \subseteq S$, $g[S']$ is T'-Connected for some unique T' such that $T \subseteq T'$ and $T' \subseteq S'$, for all $S \subseteq S'$. Equivalently, $\mathcal{G}_T^S \subseteq \bigcup_{T' \subseteq S' : T \subseteq T'} \mathcal{G}_{T'}^{S'}$ for all $S \subseteq S'$.*

Proof. The proof is straightforward and comes from the definitions of $g[S]$ and T-Connectedness. Recall that $g[S] = g - \sum_{\{i,j\} \subseteq X \setminus S} g_{ij}$. If $g[S]$ is T-Connected, by definition, there exists a $t - t'$ path in $g[S]$ for all $t, t' \in T$, $t \neq t'$. Each of these paths survives in $g[S']$ for some $S' \supseteq S$, since $g[S'] = g[S] + \sum_{i \in S'} \sum_{j \in S' \setminus S} g_{ij}$. Therefore, each $t, t' \in T$ is connected under $g[S']$.

Let T' be the largest set of nodes in S' such that each $t, t' \in T'$ is connected under $g[S']$ and $T' \supseteq T$. Clearly $T' \neq \emptyset$, since T itself is connected under $g[S']$ by the above logic. Then $g[S']$ is T'-Connected.

To show that T' is unique, suppose it isn't and consider $T'' \subseteq S'$, but $T'' \neq T$. Note that $T \subset T''$, by construction, so either i) $\exists t'' \in T''$ such that $t'' \notin T'$ or ii) $T'' \subseteq T'$. Suppose it is case (i), then T' was not the largest set of nodes in S' such that each $t, t' \in T'$ is connected under $g[S']$, since t'' is connected to some $t \in T$ (by $T \subseteq T''$) and $T' \cup \{t''\} \supseteq T'$, a contradiction. Next, suppose it is case (ii), then $g[S']$ is not T''-Connected, since $\exists t'' \in T' \setminus T''$ that is connected to some $t \in T \subseteq T''$ by construction, which is a contradiction. Therefore, T' is unique. $\square$

Proposition 1. *If $\Gamma_x(T \mid S)$ is a random consideration set mapping such that i) $\Gamma_x(T \mid S) \in \{0,1\}$ for all $x \in T \subseteq S$ and ii) Γ_x satisfies A1 - A3, then $\Gamma_x(S)$*

satisfies B1 - B3 where $\Gamma_x(S) = T$ *for* $\Gamma_x(T \mid S) = 1$.

Proof. Let $\Gamma_x(T \mid S)$ be a random consideration set mapping such that i) $\Gamma_x(T \mid S) \in \{0, 1\}$ for all $x \in T \subseteq S$ and ii) Γ_x satisfies A1 - A3. Let $\Gamma_x(S) = T$ for $\Gamma_x(T \mid S) = 1$

B1. Consider $\Gamma_x(S)$. Since $\Gamma_x(T \mid S)$ satisfies A1, $\Gamma_x(T \mid S) \leq \sum_{T' \subseteq S' : T \subseteq T'} \Gamma_x(T' \mid S')$ for each $x \in T \subseteq S$ and $S' \supseteq S$. By definition of $\Gamma_x(T' \mid S')$, which requires that $\Gamma_x(T' \mid S') = 1$ for some unique $T' \subseteq S'$, there exists some T'' such that $\Gamma_x(T'' \mid S') = 1$. This, together with $\sum_{T' \subseteq S' : T \subseteq T'} \Gamma_x(T' \mid S') \geq 1 = \Gamma_x(T \mid S)$, ensures that $T \subset T''$. Then $\Gamma_x(S) = T \subseteq T'' = \Gamma_x(S')$ and $\Gamma_x(S)$ satisfies B1.

B2. Consider $\Gamma_x(S)$ and $\Gamma_y(S)$, assuming $x, y \in S$ and that $y \in \Gamma_x(S)$. Then $\exists T, T'$ where $\Gamma_x(T \mid S) = 1 = \Gamma_y(T' \mid S)$. Since $y \in \Gamma_x(S)$, $y \in T$ and $T = T'$, since by A2 $y \in T$ implies $\Gamma_x(T \mid S) = \Gamma_y(T \mid S) = 1$. Then $\Gamma_x(S) = \Gamma_y(S)$ and $\Gamma_x(S)$ satisfies B2.

B3. Let $z \in \Gamma_x(S)$ and $z \notin \Gamma_x(S \setminus \{y\})$. Then $z \in T$ for T such that $\Gamma_x(T \mid S) = 1$ and $z \notin T'$ for $\Gamma_x(T' \mid S \setminus \{y\}) = 1$. Note that T' is the <u>only</u> subset of $S \setminus \{y\}$ for which $\Gamma_x(T' \mid S \setminus \{y\}) > 0$, by definition of $\Gamma_x(T \mid S)$ and $\Gamma_x \in \{0, 1\}$. Since $\Gamma_x(T \mid S)$ satisfies A3, $\nexists T''$ with $y \in T''$ such that $\Gamma_x(T'' \mid S \setminus \{z\}) > 0$ and $\nexists T'''$ with $z \in T'''$ such that $\Gamma_y(T'' \mid S \setminus \{x\}) > 0$. Then, by definition of $\Gamma_x(S)$, $y \notin \Gamma_x(S \setminus \{z\})$ and $z \notin \Gamma_y(S \setminus \{x\})$. Therefore, $\Gamma_x(S)$ satisfies B3.

$\square$

Proposition 2. *If a random consideration set mapping has a random network consideration set mapping representation, it satisfies RNC Symmetry, RNC Upward Monotonicity, and RNC Path Connectedness.*

Proof. Suppose a random consideration set mapping Γ has a random network consideration set mapping G:

RNC Symmetry Consider any $T \subset S$ with $\{x, y\} \subseteq T$ and $x \neq y$. If $G_x(T \mid S) > 0$, then there exists some network $g[S]$ that is T-Connected. Since $y \in T$, $g[S]$ will also be included in $G_y(T \mid S)$. Therefore, $G_x(T \mid S) \leq G_y(T \mid S)$. $G_y(T \mid S) \leq G_x(T \mid S)$ by the same logic. Finally, if $G_x(T \mid S) = 0$, then there are no $y \in \mathcal{G}$ such that $g[S]$ is T-Connected. This will hold regardless of the starting point in T, so $G_y(T \mid S) = 0$ as well. Then RNC Symmetry holds

RNC Upward Monotonicity Given Lemma 1, the proof is trivial. With $\}[\mathcal{S}]_{\mathcal{T}} \subseteq \bigcup_{T' \subseteq S' : T \subseteq T'} \mathcal{G}_{T'}^{S'}$ for all $S \subseteq S'$, the statement follows directly from the definition of $G_x(T \mid S)$.

RNC Path Connectedness Let T be such that $z \in T$ with $G_x(T \mid S) > 0$. Then there exists some $g[S] \in \mathcal{G}_T^S$ where $f(g[S]) > 0$. Since $\nexists T' \subseteq S \setminus \{y\}$ with $z \in T'$ such that $G_x(T' \mid S \setminus \{y\}) > 0$, then every path that connects x to z under $g[S]$ must include y as an intermediate node. To see why this is the case, consider some x-z path in $g[S]$ that does not include y as an intermediate node. When y is removed from S, this path remains (since y was not on this path under $g[S]$) and $G_x(T' \mid S \setminus \{y\}) > 0$ for $T' = \{j \mid j \text{ is connected to some node on this } x\text{-}z \text{ path in } g[S]\}$ since $f(g[S]) > 0$. Since there exists an x-y-z path in y^S, we can consider each sub-path independently.

Consider the x-y sub-path. When z is removed from S, this path survives, and if we let $T'' = \{j \in S \setminus \{z\} \mid j \text{ is connected to some node on the } x\text{-}z \text{ path in } g[S]\}$, then $G_x(T'' \mid S \setminus \{z\}) > 0$, since $f(g[S]) > 0$.

By similar logic, $G_y(T''' \mid S\backslash\{x\}) > 0$ for $T''' = \{j \in S\backslash\{x\} \mid j$ is connected to some node on t

□

Lemma 2. *For any RNC π, x is revealed preferred to y if $\exists S \subseteq X$ such that:*

$$\pi_y(x \mid S) > 0 \tag{A.4}$$

Proof. For some π with an RNC representation, suppose that there exists $S \subseteq X$ such that $\pi_y(x \mid S)$. Choose some RNC representation of π, call it $(\hat{F}, \hat{\succ})$. By definition of π with an RNC representation, this implies that $\exists T \subseteq S$ with $\{x, y\} \subseteq T$ where $G_y^F(T \mid S) > 0$. Further, by definition, x is $\hat{\succ}-$best in T. Since $y \in T$, $x\hat{\succ}y$ according to this representation.

To show that $x \succ y$ for all RNC representations, suppose not. Then $\exists(F', \succ'$ $) \neq (\hat{F}, \hat{\succ})$ that also represents π, but for which $y \succ' x$. However, $\pi_y(x \mid S) > 0$ implies that $x \succ' y$, by the above logic, a contradiction. □

Lemma 3. *For some RNC π, x is revealed preferred to y if, and only if, xP_Ry.*

Proof. $\rightarrow$:

The if part of this proof is trivial. Since xP_Ry, either xPy, indicating that x is revealed preferred to y directly by Lemma 2 or $\exists k$ such that xPk and kPy. In the latter case, xPk and kPy in all representations of π, again by Lemma 2, which, by transitivity of $\succ$ implies that xP_Ry in all RNC representations of π. Thus, x is revealed preferred to y.

$\leftarrow$:

Suppose that x is revealed preferred to y, but not xP_Ry. Since P_R is the transitive closure of P, it can be written as $P_R = \bigcup_{i=1,2,3,...} P^i$, where $P^1 = P$ and $P^{i+1} = P \circ P^i$. Then, if xP_Ry, $(x, y) \in P^i$ for some i and there then exists some finite sequence $\{k_0, k_1, ..., k_n\}$ where $xPk_0P...Pk_nPy$. Since $(x, y) \notin P_R$, there exists no such finite sequence.

By Lemma 2, all representations of π will be such that $P \subseteq \succ$. Select one, calling it $(F, \succ)$. From this, we construct an additional RNC $(F, \succ')$, where $\succ'$ is such that:

1. $P \subseteq \succ'$
2. $(y, x) \in \succ'$
3. $\succ'$ is transitive

Note that by the last requirement, $P_R \subseteq \succ'$ (all transitive supersets of P will include P_R). Since $(x, y) \notin P_R$, the construction of $\succ'$ to include P and (y, x) is valid.

We claim that $(F, \succ')$ also represents π. To show this, let $\pi^{\succ'}$ be RNC choice probabilities under $\succ'$ and consider $\pi_w(z \mid S)$ and $\pi_w^{\succ'}(z \mid S)$ for some arbitrary $w, z \in S$.

Case 1: $(z, w) \in P$ or $(w, z) \in P$

Suppose, to the contrary, that $\pi_w(z \mid S) > \pi_w^{\succ'}(z \mid S)$, without loss of generality. Then $\exists$ some $T \subseteq S$ with $\{w, z\} \subseteq T$ and z as $\succ$-best in T, but z is not $\succ'$ $-$best in T, with some $g[S] \in \mathcal{G}_T^S$ such that $f(g[S]) > 0$. Let $t^{\succ'}$ be the $\succ'$-best element in T. Note that $t^{\succ'} \neq w$, since $(z, w) \in P \subseteq \succ'$. Let $T' \subseteq T$ be the set of all nodes on some $w - t^{\succ'}$ path in $g[S]$. Now consider $\pi_{t^{\succ'}}(z \mid T')$. Since $f(g[S]) > 0$, $t^{\succ'}$ and z are connected under $g[S]$, and $\pi_{t^{\succ'}}(z \mid T') > 0$, since z is $\succ$-best in $T \supseteq T'$. Then $(z, t^{\succ'}) \in P \subseteq \succ'$, a contradiction.

For $(w, z) \in P$, follow the above logic, reversing the roles of w and z.

Case 2: (z, w) and $(w, z) \notin P$

Since neither (z, w) nor $(w, z) \in P$, by definition $\pi_w(z \mid S) = \pi_z(w \mid S) = 0$ for all $S \subseteq X$. Suppose under $\succ$, $z \succ w$. Then for $\pi_w(z \mid S) = 0$ to be true, either i) $f(g[S]) = 0$ for all $g[S]$ such that z and w are connected or ii) for all $g[S]$ such that $f(g[S]) > 0$ and z is connected to w, all $w - z$ paths in $g[S]$ include some node k such that $k \succ z$ and $k \succ w$. Clearly, if F such that (i) holds, $\pi_w^{\succ'}(z \mid S) = \pi_z^{\succ'}(w \mid S) = 0$.

Then if (ii) holds, for each $g[S]$ such that $f(g[S]) > 0$ and z is connected to w, and for each a $w - z$ path that includes some node k such that $k \succ z$, consider independently the $w - k$ and $k - z$ sub-paths. Denote the sets of nodes for each of these sub-paths $T_{w-k}^{g[S]}$ and $T_{k-z}^{g[S]}$, respectively. Then $\pi_w(k \mid T_{w-k}^{g[S]}) > 0$ and $\pi_z(k \mid T_{k-z}^{g[S]}) > 0$, since $f(g[S]) > 0$. It follow then, that $\{(k, w), (k, z)\} \subseteq P \subseteq \succ'$ and $\pi_w^{\succ'}(z \mid S) = \pi_z^{\succ'}(w \mid S) = 0$ for all $S \subseteq X$.

Therefore, $\pi^{\succ'} = \pi$ and $(F, \succ')$ also represents π. Since both $(F, \succ)$ and $(F, \succ')$ represent π, but $(y, x) \in \succ'$, x is not revealed preferred to y, a contradiction. $\square$

Lemma 4. *Let π be an RNC and let x and y be such that there exists some set $S \supseteq \{x, y, z\}$ such that the following holds:*

$$\pi_z(x \mid S) > \pi_z(x \mid S \setminus \{y\}) \tag{A.5}$$

Then $(x, y) \in P$ and x is revealed preferred to y.

Proof. Let x and y be such that there exists some set S and $z \in S$ such that $\pi_z(x \mid S) > \pi_z(x \mid S \setminus \{y\})$. Then when S is available, there exists some network $g[S] \in \mathcal{G}_T^S$ for some $T \supseteq \{x, z\}$ such that x is $\succ$ $-$best in T and $f(g[S]) > 0$. Suppose, to the contrary, that $y \notin T$. Then if $g[S]$ is T-Connected, $g^{S \setminus \{y\}}$ is also T-Connected. Then $\pi_z(x \mid S) = \pi_z(x \mid S \setminus \{y\})$, since this will hold for all T such that $y \notin T$ and x is $\succ$-best in T, a contradiction. Then $y \in T$.

Note that if $g[S]$ is T-Connected, g^T is T-Connected. Then $\pi_y(x \mid T) > 0$, since $f(g[S]) = f(g^T) > 0$, $\{x, y\} \subseteq T$, x is $\succ$-best in T, and g^T is T-Connected. Therefore, $(x, y) \in P$ and x is revealed preferred to y. $\square$

Lemma 5. *Let π be a PM-RNC with representations $(\mu^\pi, \succ^\pi)$. Then $\mu^i = \bar{\mu}$ for all $(\mu^i, \succ^i)$ representations of π (i.e. μ is unique).*

Proof. This lemma is fairly straightforward and is a function of the restrictions imposed by the particular consideration structure.

Note that for this lemma to hold, we must show that for any $(succ_i, \mu_{i)}$ and $(\succ_j, \mu^j)$ that both represent π, $\mu^i_{kl} = \mu^j_{kl}$ $k, l \in X$ such that $k \neq l$.

Suppose $k \succ l$ (or k is revealed preferred to l). Then the following must be true $\pi_l(k \mid \{k,l\}) = \mu^i_{kl}$ since μ^i represents π. Similarly, $\pi_l(k \mid \{k,l\}) = \mu^j_{kl}$. It follows that $\mu^i_{kl} = \mu^j_{kl}$ for all k, l such that k is revealed preferred to l.

But what if k cannot be revealed preferred to l? Assume to the contrary that $\mu^i_{kl} > 0$ for some μ^i that represents π. By definition,

$$\pi_l(k \mid \{k,l\}) = \begin{cases} \mu^i_{kl}, & \text{for } k \succ^i l \\ \\ 0, & \text{for } l \succ^i k \end{cases} \tag{A.6}$$

Since $\mu^i_{kl} > 0$ and $\not(k \succ l)$, $l \succ^i k$. By definition,

$$\pi_k(l \mid \{k,l\}) = \begin{cases} \mu^i_{kl}, & \text{for } l \succ^i k \\ \\ 0, & \text{for } k \succ^i l \end{cases} \tag{A.7}$$

Since $\mu^i_{kl} > 0$ and $l \succ^i k$, $\pi_k(l \mid \{k,l\}) > 0$, which implies that $l \succ k$, a contradiction.

Thus, for any two elements k and l where we cannot reveal preference, $\mu^i_{kl} = 0$ for any μ^i that represents π.

$\square$

Proposition 3. *If an RNC G_x has a PM-RNC representation, it satisfies PM-RNC Binary Separability.*

Proof. The proof is written for $S' = \{z\}$, but the aggregate of the logic to larger S' is trivial.

Let G_x be an RNC with a PM-RNC representation, which is denoted as the matrix of consideration weights μ. For any network g, the probability that it occurs can be written as follows:

$$f(g) = \prod_{(i,j)\in X^2} [\mathbb{1}\{g_{ij} = 1\}\mu_{ij} + \mathbb{1}\{g_{ij} = 0\}(1 - \mu_{ij})]$$

Then for any starting point x in set S, the probability that set $T \subseteq S$ is considered is given by the following, for any non-trivial probability:

$$\begin{aligned}\Gamma_x(T \mid S) &= \sum_{g\in\mathcal{G}_T^S} f(g)\\ &= \sum_{g\in\mathcal{G}_T^S} \prod_{(i,j)\in X^2} [\mathbb{1}\{g_{ij} = 1\}\mu_{ij} + \mathbb{1}\{g_{ij} = 0\}(1 - \mu_{ij})]\end{aligned}$$

Note that by Lemma 1, if $g^{S\cup\{z\}}$ is T-Connected, then $g[S]$ is also T-Connected. In other words, $\Gamma_x(T \mid S \cup \{z\})$ can be constructed by beginning with $\mathcal{G}_T^S$ and subtracting out those networks for which $g[S]$ is not T-Connected.

Consider a partition $\mathcal{P}(\mathcal{G}_T^S)$ of $\mathcal{G}_T^S$ into subsets where g and g' are included in the same subset if $g_{ij} = g'_{ij}$ for all $\{i, j\} \neq \{t, z\}$ for some $z \notin S$ and $t \in T$. Let P be an arbitrary one of these subsets. We define the following:

$$f(P) \equiv \sum_{g\in P} f(g) \tag{A.8}$$

where $f(P)$ is taken to be probability over all networks in P. Recalling that on each $g \in P$ is such that $g[S]$ is T-Connected, restrict attention only to those in P such that $g^{S\cup\{z\}}$ is T-Connected. There is clearly a single $g \in P$ such that $g^{S\cup\{z\}}$ is

T-Connected, the network g such that $g_{tz} = 0$ for all $t \in T$ (otherwise, $g^{S\cup\{z\}}$ would not be T-Connected). Let this unique $g \in P$ be denoted $g^*(P)$. Since $g^*_{tz}(P) = 0$ for all $t \in T$, it follows that $f(g^*(P)) = \prod_{t\in T}(1-\mu_{tz})f(P)$.

Then $\Gamma_x(T \mid S \cup \{z\})$ can be constructed as follows:

$$
\begin{aligned}
\Gamma_x(T \mid S \cup \{z\}) &= \sum_{g\in\mathcal{G}_T^{S\cup\{z\}}} f(g) \\
&= \sum_{P\in\mathcal{P}(\mathcal{G}_T^S)} f(g^*(P)) \\
&= \sum_{P\in\mathcal{P}(\mathcal{G}_T^S)} \prod_{t\in T}(1-\mu_{tz})f(P) \\
&= \sum_{P\in\mathcal{P}(\mathcal{G}_T^S)} \prod_{t\in T}(1-\mu_{tz}) \sum_{g\in P} f(g) \\
&= \prod_{t\in T}(1-\mu_{tz})\Gamma_x(T \mid S)
\end{aligned}
$$

The result then directly follows an observation that $(1-\mu_{tz})$ is simply equal to $\Gamma_z(\{z\} \mid \{t,z\})$ in the PM-RNC model. $\square$

Proposition 4. *In the RNC Advertising Game with parameters α, β, and c, such that (α, β, c) is in the unit cube, there exist non-empty subsets of the parameter space where:*

1. *$\sigma^* = (1,0,1)$ is supported as a Nash Equilibrium*

2. *$\sigma' = (0,1,0)$ is not supported as a Nash Equilibrium*

3. *Aggregate profit under σ' is higher than under σ^**

Proof. 1. $\sigma^* = (1, 0, 1)$ **is supported as a Nash Equilibrium**:

First, consider each firm's incentives under $\sigma^* = (1, 0, 1)$. Firm A's expected profit under σ^* is as follows:

$$\pi_A(\sigma^*) = \frac{3\alpha}{4} + \frac{1-2\alpha}{3} - c$$

In order for Firm A to not have an incentive to unilaterally deviate to $\sigma_A = 0$, the following condition must then hold:

$$\begin{aligned} \frac{3\alpha}{4} + \frac{1-2\alpha}{3} - c &> \frac{1-2\alpha}{3} \\ \frac{3\alpha}{4} &> c \end{aligned} \tag{A.9}$$

The condition for Firm C is identical.

Conditional on Firm's A and C choosing to advertise, Firm B has no incentive to advertise if the following holds:

$$\begin{aligned} \frac{2\alpha}{3} + \frac{1-2\alpha}{3} - c &< \frac{\alpha}{2} + \frac{1-2\alpha}{3} \\ \frac{\alpha}{6} &< c \end{aligned} \tag{A.10}$$

Clearly there exist positive α and c such that conditions A.9 and A.10 hold.

2. $\sigma' = (0, 1, 0)$ **is not supported as a Nash Equilibrium**:

   For Firm A, there exists an incentive to unilaterally deviate from $\sigma'_A = 1$ if the following holds:

$$\begin{aligned} \alpha + \frac{1-2\alpha}{3} - c &> \frac{\alpha}{2} + \frac{1-2\alpha}{3} \\ \frac{\alpha}{2} &> c \end{aligned} \tag{A.11}$$

   The condition is identical for Firm C.

   Clearly, for any feasible and strictly positive α, A.9 - A.11 are satisfied for any $c \in (\frac{\alpha}{2}, \frac{3\alpha}{4})$.

3. **Aggregate profit under σ' is higher than under σ^***

   This should be clear from the definition of the profit function. Under σ^*, aggregate profit is equal to $1 - 2c$, since two firms are advertising, whereas under σ' it is simply $1 - c$.

$\square$

A.7 Additional Results

A.7.1 Results by Baseline/Context

Table A.13: Correct Rate by Treatment

	NC	C
Mean	0.851	0.863
Std Error	0.009	0.009
N	1733	1555

Wilcox $p > 0.10$ for $H_0 : \mu_C = \mu_{NC}$.

Table A.14: Monotonicity Violations by Context

	Baseline	Context
Mean	0.791	0.805
Std Error	0.012	0.013
N	1140	1000

Wilcox $p > 0.10$ for $H_0 : \mu_{Baseline} = \mu_{Context}$

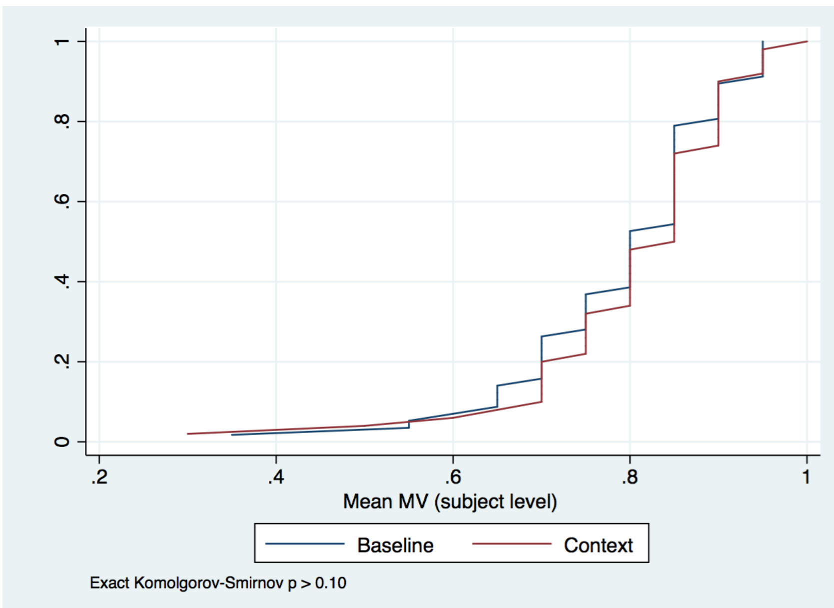


Figure A.8: Cumulative Distribution of Mean MV by Subject and Treatment

Table A.15: Path Connectedness by Context

	Baseline	Context
Mean	0.456	0.452
Std Error	0.067	0.078
N	57	42

Mann-Whitney $p > 0.10$ for $H_0 : \mu_{Baseline} = \mu_{Context}$

A.7.2 Omitted Results

Table A.16: Determinants of Consideration Set Size

	Model 1	Model 2	Model 3
N	0.217***	0.218***	0.204***
	(0.011)	(0.011)	(0.012)
Period	0.0271***	0.0264***	0.0189**
	(0.008)	(0.008)	(0.009)
Female		0.146	0.529
		(0.466)	(0.512)
$\hat{\mu}_{GPA}$		2.131	
		(1.418)	
Cognitive Score			-0.450
			(1.137)
Constant	4.241***	2.687**	4.520***
	(0.234)	(1.106)	(0.859)
Observations	25811	25811	21525

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.8 Instructions

A.8.1 Baseline

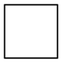
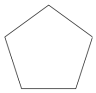
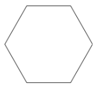
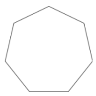
Instructions

Thank you for participating in the experiment today. At this time, please be sure that your cell phone is turned off and stored away. At no point during this experiment should you use your cell phone or any other electronic device. Also, please refrain from communicating with any other subject in the lab today. Failure to follow these rules will result in your expulsion from the lab and you will forfeit any cash earnings you may have otherwise received.

This is an experiment in decision making. You will be paid a $7 guaranteed show-up fee in addition to earnings based on your decisions in the experiment.

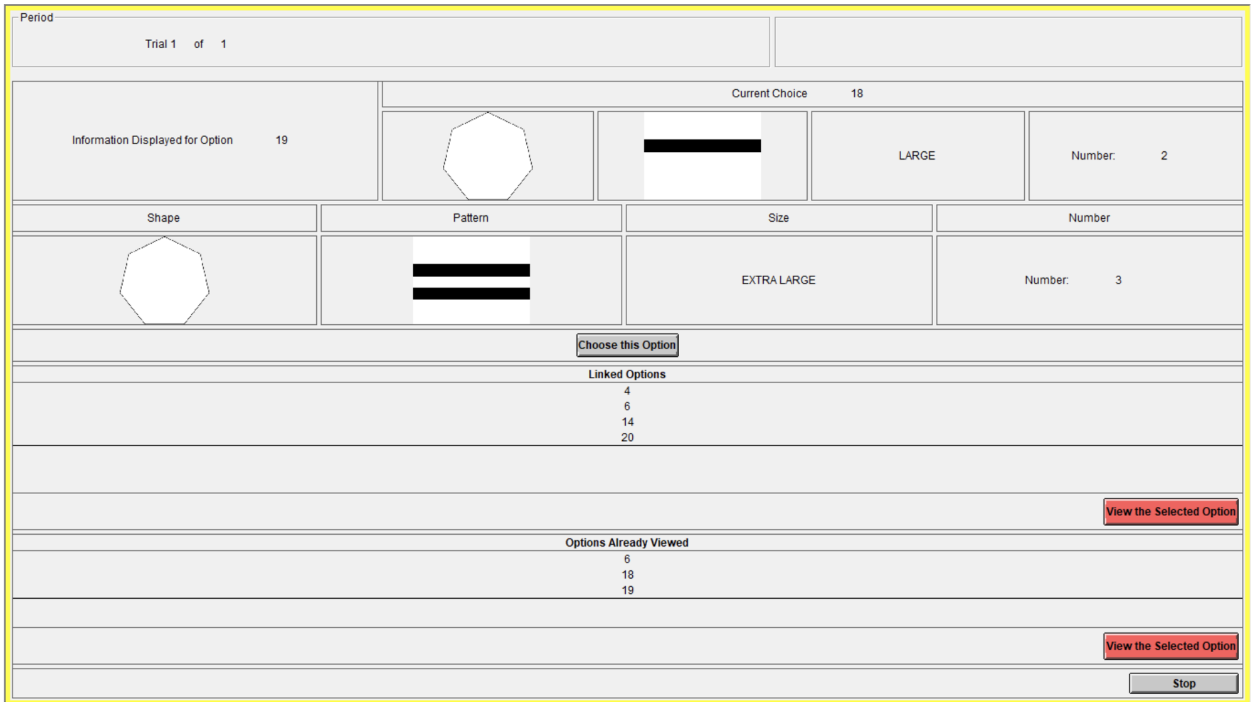
Decision Environment

In each of 31 periods, you will be faced with a number of options from which you can select one. Each option has 4 attributes: Shape, Pattern, Size, and Number. The value of an individual attribute is given in Experimental Currency Units (ECUs) in the following table:

Shape	Value	Pattern	Value	Size	Value	Number	Value
	1 ECU		1 ECU	EXTRA SMALL	1 ECU	1	1 ECU
	2 ECU		2 ECU	SMALL	2 ECU	2	2 ECU
	3 ECU		3 ECU	MEDIUM	3 ECU	3	3 ECU
	4 ECU		4 ECU	LARGE	4 ECU	4	4 ECU
	5 ECU		5 ECU	EXTRA LARGE	5 ECU	5	5 ECU

The value of a given option is the sum of the value of its attributes as per the table above.

In each period, you will be shown a version of the following screen:



The screen is composed of the following parts (left to right, top to bottom):

Option Label: this is the option for which information is currently displayed. In the example, Option 19 is shown along with information on the Attributes of Option 19. Option Labels have been chosen randomly for each Period and do not reflect the value of the option. Moreover, two options with the same option label may have different values in different periods.

Current Choice: this is the option that you are currently holding as your choice. This will be explained in detail below. In addition to the label for your Current Choice, you are shown information about the Attributes for your Current Choice for your reference.

Attribute Information: these are the attributes for the option currently displayed. The value of each option is the sum of its attributes as according to the table above. For example, the value of Option 19 in the example above is 16 ECU

(5 for Heptagon + 3 for Two-Bar Pattern + 5 for Extra Large + 3 for Number 3 = 16).

Choose this Option Button: you can click this button to change your Current Choice to the option that is currently displayed. If your Current Choice is the option that is currently displayed, this portion of the screen with display The option currently displayed is your Current Choice.

Linked Options: this is a clickable list of options that are Linked to the currently displayed option. An option is Linked to the currently displayed option if it shares 2 or more attributes with the currently displayed option. For example, if there was another Option in the current Period with a Two-Bar Pattern and the Number 3, it would be shown in the list of Linked Options for Option 19 in the screenshot above. Note that this list may be quite long, in which case you will see a scroll bar next to the list of Linked Options.

Options Already Viewed: this is a clickable list of options that have already been viewed by you in the given period. You can click on any option in this list and click View Selected Option to view information for that option again. Again, if the list of Options Already Viewed gets sufficiently long, you will be shown a scroll bar next to the list. **Note**: you can only view information for options other than the one currently displayed by either clicking on it in the Linked Options menu or the Options Already Viewed menu.

Stop: if you would like to Stop looking at information for the available options and would not like to change your Current Choice, you can click the Stop button.

Period Duration

In each period, you will have up to 75 seconds to evaluate all of the available options and make choices. At any time, you can click Stop and you will not be shown any more information on any of the options for the given period. Note: in order to move on to the next period, you must wait for the entire 75 seconds to pass in the current period. Thus, if you Stop after, say, 45 seconds in the current period, you will still have to wait the remaining 30 seconds for the period to end in order to move on to the next period.

Choices

At the end of each Period, a random time between 2 and 75 seconds will be chosen and your Current Choice held at that time will be implemented as your chosen option for that Period. Each time between 2 and 75 seconds is equally likely to be chosen.When evaluating options during the 75 seconds, you will not know at what time your Current Choice will be implemented as your chosen option for that Period. At the beginning of each Period, you will start off being shown information for one particular Option but will have no Current Choice. If you do not have a Current Choice at the time chosen randomly by the computer program to implement your choice, you will be paid nothing for the current Period. Thus, it is in your best interest to choose any Option as soon as possible. You can then replace it with a better option when/if you find an option that has a higher value.

At the end of each period, you will be told i) at what time your Current Choice was implemented, ii) which Option you held at that time, and iii) what the value (in ECU) of that option was.

For clarification purposes, consider the following example: a subject is in

Time	0 Seconds	30 Seconds	40 Seconds
Option	Option 1	Option 2	Option 3
Value	10 ECU	12 ECU	14 ECU

Period 4 with a time limit of 75 seconds and they immediately choose the first option shown to them, Option 1, which has a value that they have determined to be 10 ECU. After 30 seconds, the subject changes their Current Choice to Option 2 with a value of 12 ECU and continues evaluating the available options. After 10 more seconds (40 in total since the start of the period), the subject selects a new Current Choice of Option 3 with a value of 14 ECU. The subject makes no further choices and the 75 seconds runs out. The subject's choices are shown in the following table:

If the time chosen randomly by the system at the end of the Period is anywhere between 2 and 30 seconds, the subject will be paid 10 ECU (for holding Option 1). If it is anywhere between 30 and 40 seconds, they will be paid 12 ECU (for holding Option 2). If it is 40 seconds or higher, they will be paid 14 ECU.

Earnings

You will be paid a guaranteed show-up fee of $7 in addition to your earnings for your decision. The value of the option that is treated as your final choice for a period (i.e. the option held by you at the time chosen by the system) is the sum of the value of its Attributes as given in the table above. Experimental Currency Units (ECUs) will be converted to cash (USD) at a rate of 1 ECU = $1 USD.

Though you will make decisions in each of 31 periods, you will only be paid for 1 of these periods. Which period will be paid will be chosen at random at the end of the experiment, with each period being equally likely to be chosen. Thus, it is in your best interest to behave in each period as if it is the period for which you will be paid.

A.8.2 Environment with Context

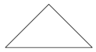
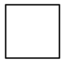
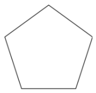
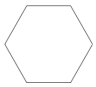
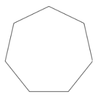
Instructions

Thank you for participating in the experiment today. At this time, please be sure that your cell phone is turned off and stored away. At no point during this experiment should you use your cell phone or any other electronic device. Also, please refrain from communicating with any other subject in the lab today. Failure to follow these rules will result in your expulsion from the lab and you will forfeit any cash earnings you may have otherwise received.

This is an experiment in decision making. You will be paid a $7 guaranteed show-up fee in addition to earnings based on your decisions in the experiment.

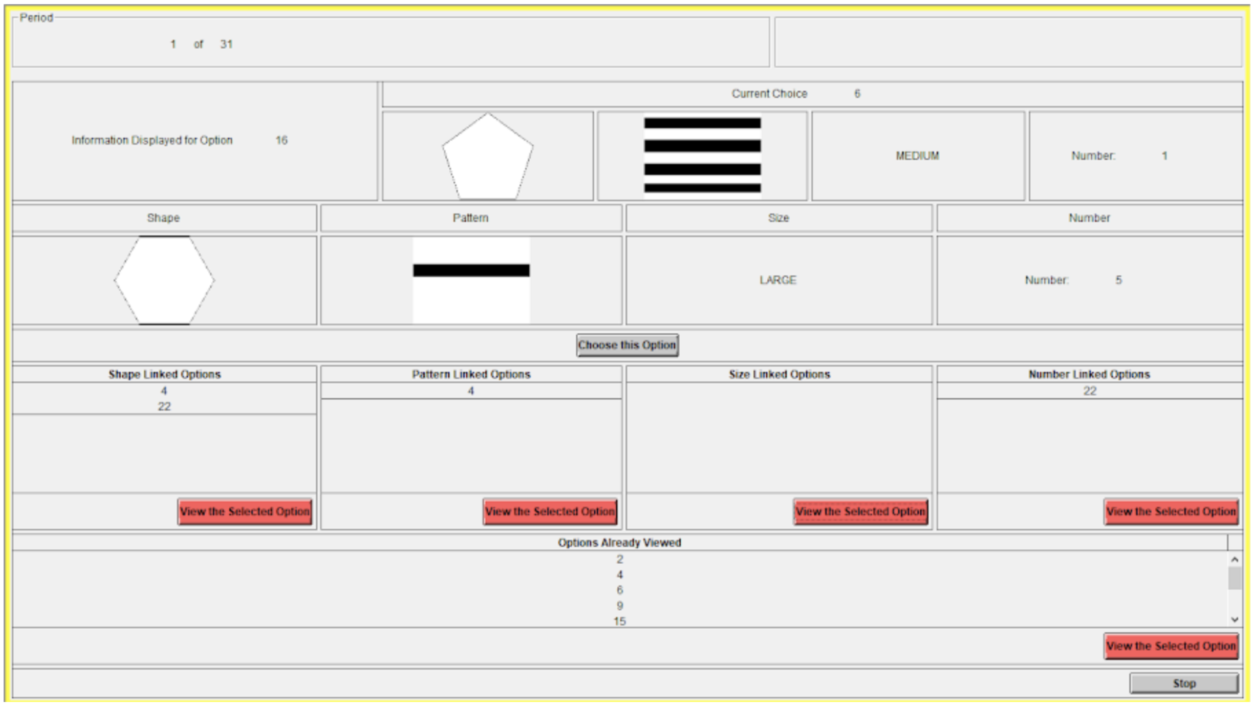
Decision Environment

In each of 31 periods, you will be faced with a number of options from which you can select one. Each option has 4 attributes: Shape, Pattern, Size, and Number. The value of an individual attribute is given in Experimental Currency Units (ECUs) in the following table:

Shape	Value	Pattern	Value	Size	Value	Number	Value
	1 ECU		1 ECU	EXTRA SMALL	1 ECU	1	1 ECU
	2 ECU		2 ECU	SMALL	2 ECU	2	2 ECU
	3 ECU		3 ECU	MEDIUM	3 ECU	3	3 ECU
	4 ECU		4 ECU	LARGE	4 ECU	4	4 ECU
	5 ECU		5 ECU	EXTRA LARGE	5 ECU	5	5 ECU

The value of a given option is the sum of the value of its attributes as per the table above.

In each period, you will be shown a version of the following screen:



The screen is composed of the following parts (left to right, top to bottom):

Option Label: this is the option for which information is currently displayed. In the example, Option 16 is shown along with information on the Attributes of Option 16. Option Labels have been chosen randomly for each Period and do not reflect the value of the option. Moreover, two options with the same option label may have different values in different periods.

Current Choice: this is the option that you are currently holding as your choice. This will be explained in detail below. In addition to the label for your Current Choice, you are shown information about the Attributes for your Current Choice for your reference.

Attribute Information: these are the attributes for the option currently displayed. The value of each option is the sum of its attributes as according to the table above. For example, the value of Option 16 in the example above is 15 ECU

(4 for Hexagon + 2 for One-Bar Pattern + 4 for Large + 5 for Number 5 = 15).

Choose this Option Button: you can click this button to change your Current Choice to the option that is currently displayed. If your Current Choice is the option that is currently displayed, this portion of the screen with display The option currently displayed is your Current Choice.

Linked Options: there are four clickable lists of options that are Linked to the currently displayed option. An option is Linked to the currently displayed option if it shares 2 or more attributes with the currently displayed option. For example, if there was another Option in the current Period with a One-Bar Pattern and the Number 5, it would be shown in a list of Linked Options for Option 16 in the screenshot above. Note that a list may be quite long, in which case you will see a scroll bar next to the list of Linked Options. The full list of Linked options is separated into four different fields, one for each Attribute: Shape, Pattern, Size, and Number. An option will be displayed in the relevant field if it meets two criteria: i) the option shares at least two Attributes with the currently displayed option and ii) it shares the Attribute for the relevant field with the currently displayed option.

For example, Option 4 also has the Hexagon Shape Attribute and the One-Bar Pattern Attribute. Since Option 16 (the currently displayed option) has both of these Attributes, Option 4 is linked to Option 16. Since it has the same Shape as Option 16, it will be listed in the Shape Linked Options list. Since it has the same Pattern as Option 16, it will also be listed in the Pattern Linked Options list. Consider, for example, another option, call it Option 12 (not displayed in the above screenshot). It has the Attributes: Square, One-Bar, Small, 4. Notice that it shares the Pattern Attribute with Option 16 (the currently displayed option): both have the Pattern One-Bar. But it does not share any other Attributes with Option 16. Therefore, it will not show up in any of the link lists when Option 16 is the currently displayed option. Especially note that it will not show up in the Pattern Linked Options list for Option 16, even though they have the same Pattern Attribute, because it does not share two or more Attributes with Option 16.

Options Already Viewed: this is a clickable list of options that have already been viewed by you in the given period. You can click on any option in this list and click View Selected Option to view information for that option again. Again, if the list of Options Already Viewed gets sufficiently long, you will be shown a scroll bar next to the list. **Note**: you can only view information for options other than the one currently displayed by either clicking on it in one of the Linked Options menus or the Options Already Viewed menu and clicking the View the Selected Option button for that list. Whenever you click on a new option from one of these lists and click View the Selection Option, all of the information on the screen (Option Label, Attribute Information, Linked Options, Options Already Viewed) will update to display information for the option to which you are navigating. The Current Choice information in the upper right of the screen will only ever change if you change your Current Choice (by choosing a new option using the Choose this Option button).

Stop: if you would like to Stop looking at information for the available options and would not like to change your Current Choice, you can click the Stop button.

Period Duration In each period, you will have up to 75 seconds to evaluate all of the available options and make choices. At any time, you can click Stop and you will not be shown any more information on any of the options for the given period. Note: in order to move on to the next period, you must wait for the entire 75 seconds to pass in the current period. Thus, if you Stop after, say, 45 seconds in the current period, you will still have to wait the remaining 30 seconds for the period to end in order to move on to the next period.

Choices

At the end of each Period, a random time between 2 and 75 seconds will be chosen and your Current Choice held at that time will be implemented as your chosen option for that Period. Each time between 2 and 75 seconds is equally likely to be chosen.When evaluating options during the 75 seconds, you will not know at what time your Current Choice will be implemented as your chosen option for that Period. At the beginning of each Period, you will start off being shown information for one particular Option but will have no Current Choice. If you do not have a Current Choice at the time chosen randomly by the computer program to implement your choice, you will be paid nothing for the current Period. Thus, it is in your best interest to choose any Option as soon as possible. You can then replace it with a better option when/if you find an option that has a higher value.

At the end of each period, you will be told i) at what time your Current Choice was implemented, ii) which Option you held at that time, and iii) what the value (in ECU) of that option was.

For clarification purposes, consider the following example: a subject is in Period 4 with a time limit of 75 seconds and they immediately choose the first option shown to them, Option 1, which has a value that they have determined to

Time	0 Seconds	30 Seconds	40 Seconds
Option	Option 1	Option 2	Option 3
Value	10 ECU	12 ECU	14 ECU

be 10 ECU. After 30 seconds, the subject changes their Current Choice to Option 2 with a value of 12 ECU and continues evaluating the available options. After 10 more seconds (40 in total since the start of the period), the subject selects a new Current Choice of Option 3 with a value of 14 ECU. The subject makes no further choices and the 75 seconds runs out. The subjects choices are shown in the following table:

If the time chosen randomly by the system at the end of the Period is anywhere between 2 and 30 seconds, the subject will be paid 10 ECU (for holding Option 1). If it is anywhere between 30 and 40 seconds, they will be paid 12 ECU (for holding Option 2). If it is 40 seconds or higher, they will be paid 14 ECU.

Earnings

You will be paid a guaranteed show-up fee of $7 in addition to your earnings for your decision. The value of the option that is treated as your final choice for a period (i.e. the option held by you at the time chosen by the system) is the sum of the value of its Attributes as given in the table above. Experimental Currency Units (ECUs) will be converted to cash (USD) at a rate of 1 ECU = $1 USD.

Though you will make decisions in each of 31 periods, you will only be paid for 1 of these periods. Which period will be paid will be chosen at random at the end of the experiment, with each period being equally likely to be chosen. Thus, it is in your best interest to behave in each period as if it is the period for which you will be paid.

C.9 Instructions

Part I

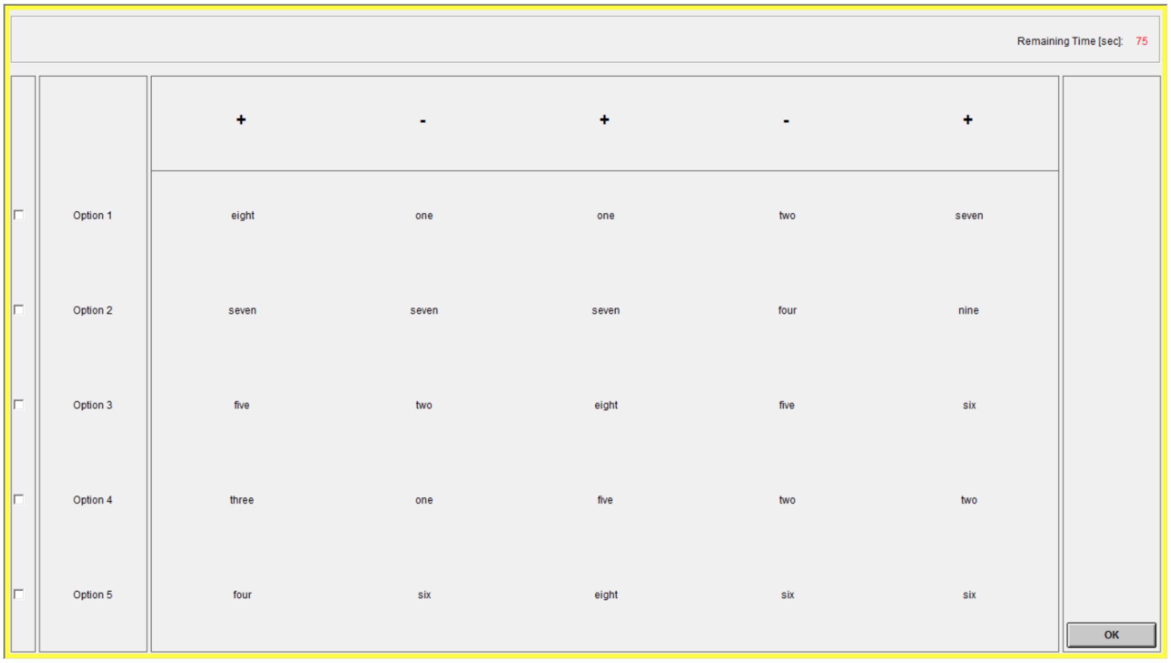
Thank you for participating in this experiment. In this session you will work alone and are not permitted to talk with any other participant. At this time, please be sure that your cell phone is turned off. At no point during the experiment are you permitted to use your cell phone or any other personal electronic device.

The Experiment

The experiment today is broken into two parts. These are the instructions for Part I of the experiment. At the conclusion of Part I, the experimenter will hand out and read instructions for Part II before proceeding. Your earnings in Part I and Part II are independent.

This is an experiment on decision-making. In each of 40 periods, you will be asked to choose one from among a number of options. You will have at most 1 minute and 15 seconds (or 75 seconds) to make this decision in each period. Each option is described by a number of attributes. Attributes take on the numbers 1-9 with each number being equally likely to be shown. The value of each option is the result of the addition and/or subtraction of these attributes and is measured in Experimental Currency Units (or ECU). The exchange rate will be as follows: 1 USD = 10 ECU. You will know whether to add or subtract each attribute based on column headers in the displayed data. While calculating these values, you will not be permitted to use a calculator or pen and paper.

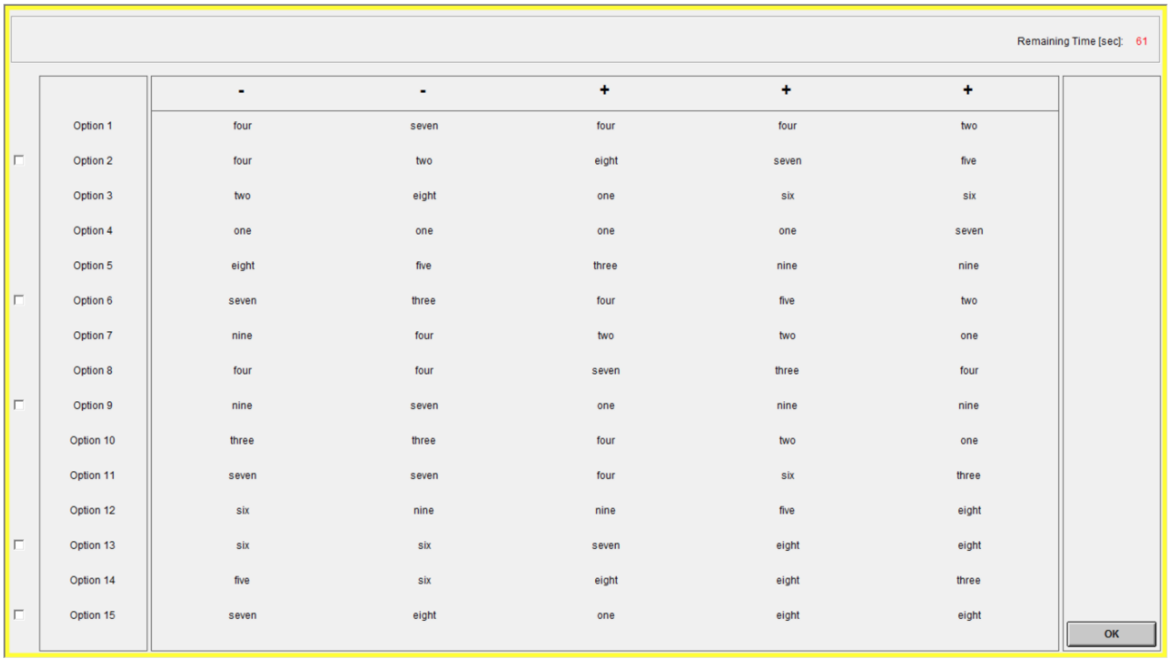
In each period, you will see a screen that looks similar to the one below:



Notice that Option 1 is accompanied by 5 numbers (shown in words) in a grid to its right. The value of Option 1 is simply the result of adding or subtracting the numbers in its corresponding row. You will know whether to add a number or subtract it based on the **plus** or **minus** sign in the column header row. Thus, the value of Option 1 is 13 ECU (or eight - one + one - two + seven = ECU). The values of Options 2-5 can be calculated in a similar way.

Variations

In each of the 40 periods, the number of available options is the same (5). However, the number of displayed options will vary. In other words, there may be some options displayed on your screen that you will not be able to select. Consider the following example:

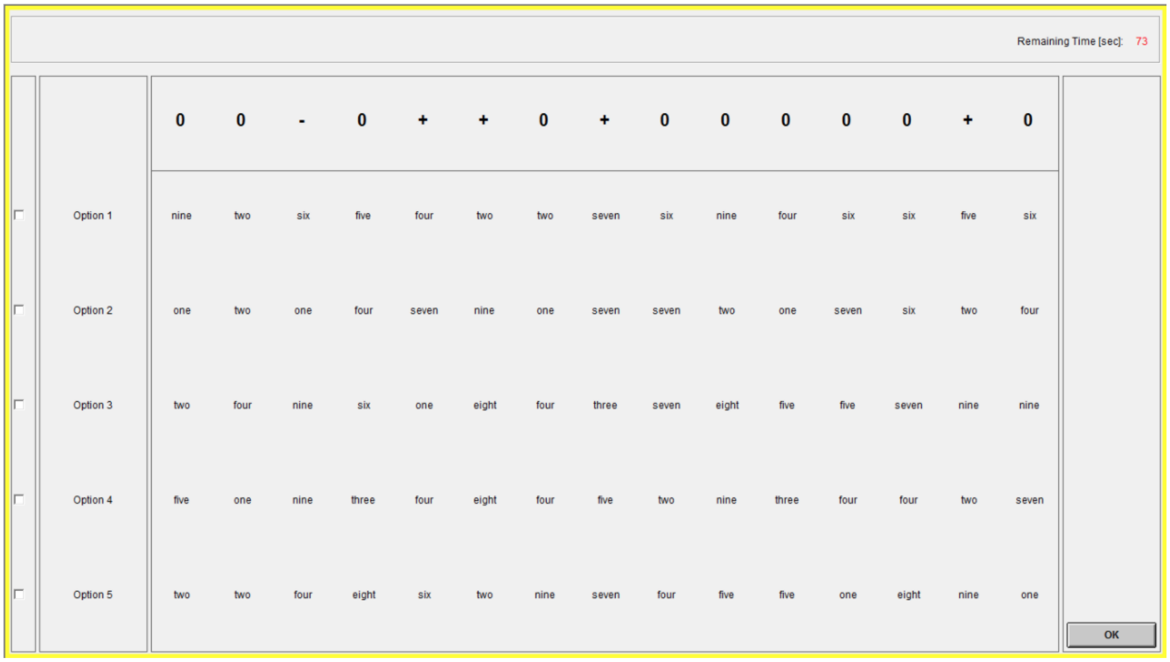

Remaining Time [sec]: 61

	-	-	+	+	+
Option 1	four	seven	four	four	two
Option 2	four	two	eight	seven	five
Option 3	two	eight	one	six	six
Option 4	one	one	one	one	seven
Option 5	eight	five	three	nine	nine
Option 6	seven	three	four	five	two
Option 7	nine	four	two	two	one
Option 8	four	four	seven	three	four
Option 9	nine	seven	one	nine	nine
Option 10	three	three	four	two	one
Option 11	seven	seven	four	six	three
Option 12	six	nine	nine	five	eight
Option 13	six	six	seven	eight	eight
Option 14	five	six	eight	eight	three
Option 15	seven	eight	one	eight	eight

OK

Note that each option still has 5 attributes in the grid. However, now Option 1 cannot be selected (this can be seen from the absence of a checkbox to the left of "Option 1). You may only select one from the following: Option 2, Option 6, Option 9, Option 13, or Option 15. Which options are available will vary between periods. Also note that the value of each option is calculated as in the first example. For example, the value of Option 2 is 14 ECU (or - four - two + eight + seven + five = 14 ECU).

In each of the 40 periods, the number of attributes per option will vary. However, in some periods, some of these attributes may be multiplied by **zeros** instead of being added or subtracted when calculating the value of each option. Consider the following example:



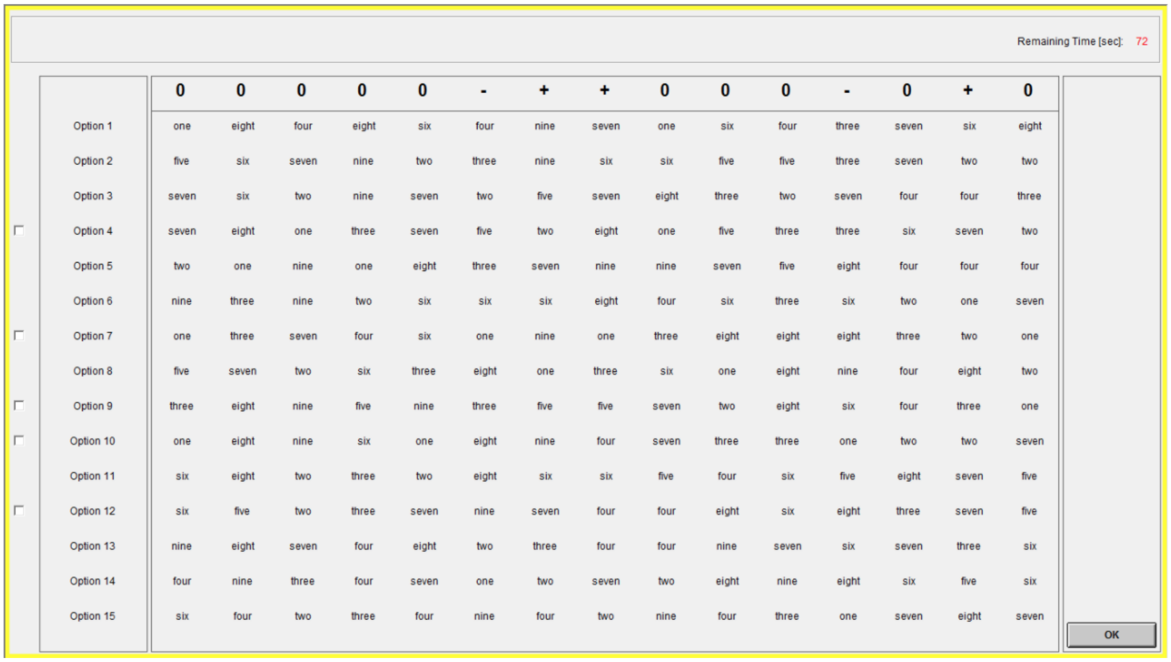
Remaining Time [sec]: 73

	0	0	-	0	+	+	0	+	0	0	0	0	0	+	0
Option 1	nine	two	six	five	four	two	two	seven	six	nine	four	six	six	five	six
Option 2	one	two	one	four	seven	nine	one	seven	seven	two	one	seven	six	two	four
Option 3	two	four	nine	six	one	eight	four	three	seven	eight	five	five	seven	nine	nine
Option 4	five	one	nine	three	four	eight	four	five	two	nine	three	four	four	two	seven
Option 5	two	two	four	eight	six	two	nine	seven	four	five	five	one	eight	nine	one

OK

Note that all displayed options are available (you can see this from the check-box to the left of each option label). However, there are additional attributes for each option (now there are 15). In contrast to the previous examples, some of these attributes are now multiplied by 0 instead of being added or subtracted when determining the value of each option. This can be seen from the zeros in the column header. For example, the value of Option 1 is 12 ECU (-six + four + two + seven + five = 12 ECU). Notice that in this calculation, the first and second attributes (nine and two) were not included because they have a 0 in the column header. The same is true for any value for which there is a zero in the column header. Which attributes have zeros (and pluses or minuses) will vary by period.

Finally, in some periods there will be additional attributes and unavailable options. Consider the following example:



Remaining Time [sec]: 72

	0	0	0	0	0	-	+	+	0	0	0	-	0	+	0
Option 1	one	eight	four	eight	six	four	nine	seven	one	six	four	three	seven	six	eight
Option 2	five	six	seven	nine	two	three	nine	six	six	five	five	three	seven	two	two
Option 3	seven	six	two	nine	seven	two	five	seven	eight	three	two	seven	four	four	three
Option 4	seven	eight	one	three	seven	five	two	eight	one	five	three	three	six	seven	two
Option 5	two	one	nine	one	eight	three	seven	nine	nine	seven	five	eight	four	four	four
Option 6	nine	three	nine	two	six	six	six	eight	four	six	three	six	two	one	seven
Option 7	one	three	seven	four	six	one	nine	one	three	eight	eight	eight	three	two	one
Option 8	five	seven	two	six	three	eight	one	three	six	one	eight	nine	four	eight	two
Option 9	three	eight	nine	five	nine	three	five	five	seven	two	eight	six	four	three	one
Option 10	one	eight	nine	six	one	eight	nine	four	seven	three	three	one	two	two	seven
Option 11	six	eight	two	three	two	eight	six	six	five	four	six	five	eight	seven	five
Option 12	six	five	two	three	seven	nine	seven	four	four	eight	six	eight	three	seven	five
Option 13	nine	eight	seven	four	eight	two	three	four	four	nine	seven	six	seven	three	six
Option 14	four	nine	three	four	seven	one	two	seven	two	eight	nine	eight	six	five	six
Option 15	six	four	two	three	four	nine	four	two	nine	four	three	one	seven	eight	seven

OK

Note that Option 1 is **unavailable** (you can see this from the absence of any checkbox to its left). Also note that there are several columns with **zeros** in the column header. The value of Option 4 is 9 ECU (-five + two + eight -three + seven= 9 ECU). Notice that the 1st through 5th attributes were not included for Option 4 (seven, eight, one, three, and seven) since these have zeros in the column header. The same is true for any column of attributes for which there is a zero in the column header. Again, which columns have zeros (and pluses/minuses) and which options are unavailable will vary by period.

Time Limit

In each period, you have 1 minute and 15 seconds (75 seconds) to submit your choice of option. You must submit your option by checking the checkbox to its left and clicking the OK button at the bottom right of the screen. If you **do not** submit your selection by clicking the OK button prior to the end of the period (i.e. within 75 seconds of the period starting), your selection will not be submitted and **you will be paid nothing** for that period. Only by selecting an option and clicking OK prior to the end of the period will your choice be submitted for the period.

Earnings

In each period, your per-period payoff is simply the value of the option you have chosen. In each of these periods, the values for each option have been chosen so that despite being the sum of both positive and negative numbers, the **value of each available option is positive**. That is, no matter which option you choose, money will never be taken away from you. 10 periods will be chosen at random and your cash earnings will be the sum of the per-period payoffs for these 10 periods, converted to US Dollars. The exchange rate will be as follows: $1 USD = 10 ECU. Your total cash earnings will be added to your show-up fee of $7.00 and your earnings from Part II of this experiment.

You will be paid your earnings privately in cash before you leave the lab.

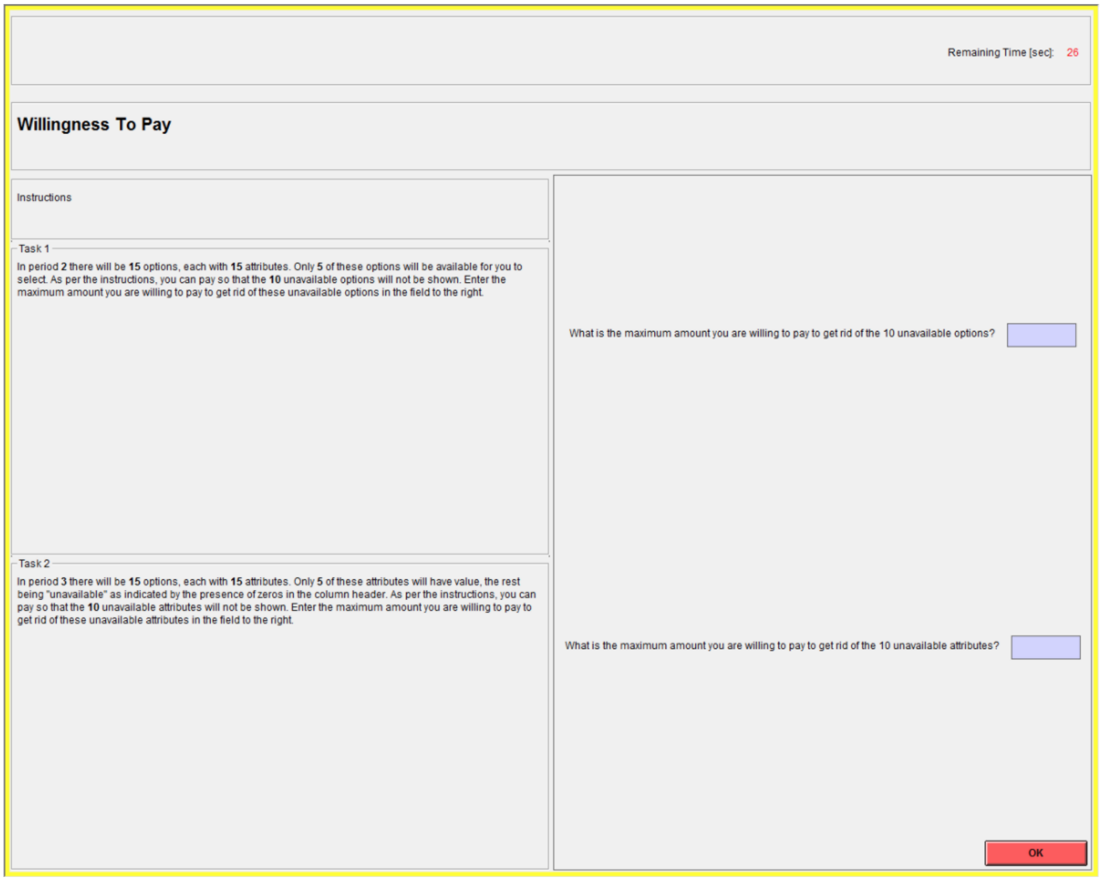
Part II

Thank you for participating in Part II of the experiment.

You will be faced with 3 periods in which you make decisions: **1** period in which you will be asked to submit two numbers (explained in detail below), and **2** periods of decision environments where you will choose from among a number of options, each described by a number of attributes. Some of these options will be unavailable for you to select and some of the attributes will not have value (as indicated by the presence of a zero in the header row). However, you will have the opportunity to pay some amount (in ECU) to get rid of these unavailable options and attributes.

In period 1, you will be asked to complete **two tasks** which will affect what you see in periods 2 and 3: **Task 1** is to enter the maximum amount you are willing to pay (in ECU) to get rid of the unavailable options to be presented in period 2, and **Task 2** is to enter the maximum amount you are willing to pay to get rid of the attributes that have no value (as indicated by the zeros in the column header; these will be referred to as unavailable attributes for the remainder of the instructions) to be presented in period 3. Note that decisions in each task will correspond to outcomes in two separate subsequent periods: Task 1 affects what you see in period 2 and Task 2 affects what you see in period 3.

The screenshot below displays what this environment will look like in period 1:



For **Task 1** and **Task 2**, two random numbers will be drawn from 0 ECU to 15 ECU. These two numbers may not be the same. These will be the selling prices for getting rid of the unavailable options or unavailable attributes, respectively. If the maximum amount you are willing to pay to get rid of unavailable options that you entered for Task 1 is above the selling price for Task 1, you pay the selling price and you will not see these unavailable options in period 2. If the maximum amount you are willing to pay to get rid of unavailable attributes is higher than the selling price for Task 2, you pay the selling price and you will not see these unavailable attributes in period 3. However, if either (or both) of the selling prices are above the maximum amount you are willing to pay, entered in period 1 for Task 1 and

Task 2, you pay nothing and the unavailable options or unavailable attributes will be shown in the respective period.

Note that you enter both of these numbers indicating your maximum willingness to pay to simplify the environments at the same time and **before** you know the result of either random number draw. That is, when you enter the maximum amount you are willing to pay to get rid of unavailable options, you will not know whether you have been able to get rid of unavailable attributes, and when you enter the maximum amount you are willing to pay to get rid of unavailable attributes, you will not know whether you have been able to get rid of the unavailable options. Also note that it is in your best interest not to overstate (or understate) the maximum amount you are willing to pay in either Task 1 or Task 2. Suppose you are willing to pay at most 5 ECU to get rid of either unavailable options or attributes. If the random is drawn and you enter exactly 5 ECU, there are two potential outcomes: either the number is higher than 5, in which case you pay nothing and the unavailable options or attributes will be displayed in the respective period, or the number is less than 5 , say 4 ECU. In this case, you pay the 4 ECU and the unavailable options or attributes are not shown. Note that you were willing to pay **at most** 5 ECU, but only had to pay 4 ECU.

Suppose instead that you overstate this amount in either Task 1 or Task 2 by entering, say, 6 ECU. Then it could be the case that the number drawn is 5.5, for example, which is less than 6 (which you have entered) but greater than 5, the true maximum amount that you are willing to pay. Because you have entered 6, you will pay the drawn amount, 5.5 ECU, which is more than you originally were willing to pay - you will have gotten rid of unavailable options or attributes, but paid more than the maximum amount you were willing to pay. On the other hand, suppose you understate this amount by entering 4 ECU. Then if the random number drawn is, say, 4.5 ECU, you will not be able to get rid of the unavailable options or attributes, but would be willing to pay this amount. Only by entering the actual maximum amount you are willing to pay in Task 1 and Task 2 will you both a) prevent having to pay more than this amount (by overstating) and b) prevent missing out on paying a lesser amount when it is profitable to do so (by understating).

Decision Environments

These decision environments will appear exactly as you have seen them in Part 1. Again, you will have 75 seconds to submit your decision. If you do not submit your chosen option by that time, no option will be submitted and you will be paid nothing for that period.

By default, in period 2 there will be 15 options, each with 15 attributes. Only 5 of these options will be available for you to select and only 5 of these attributes will have value (as indicated by the presence of a + or - in the column header). You can pay to have the **10 unavailable options** not displayed in this period. No matter what, each of the displayed options will have 15 attributes, 10 of which will have zeros in the column header. Whether the 10 unavailable options are displayed depends on the result of your choice in Task 1, described in detail above.

By default, in period 3 there will be 15 options, each with 15 attributes. Only

5 of these attributes will have value - the rest are unavailable (as indicated by the presence of zeros in the column header) and only 5 of these options will be available for you to select. You can pay to have the **10 unavailable attributes** not displayed in this period. No matter what, there will be 15 options displayed (5 of which will be available for selection). Whether the 10 unavailable attributes are displayed depends on the result of your choice in Task 2, described in detail above.

Payoff Calculation

In each of periods 2 and 3, your per-period payoff is simply the value of the option you have chosen. In each of these periods, the values for each option have been chosen so that despite being the sum of both positive and negative numbers, the **value of each available option is positive**. That is, no matter which option you choose, money will never be taken away from you.

Choices in all periods contribute to your payoffs for this part of the experiment. In the first period, if you are able to get rid of either unavailable options or attributes or both, the relevant random number that was drawn is subtracted from your payoffs. In each of the decision periods, the value of the option you have chosen will be added to your payoffs, with the value of each option calculated as in Part I of this experiment. The exchange rate will be as follows: \$1 USD = 10 ECU. Your total cash earnings will be added to your show-up fee of \$7.00 and your earnings from Part I of this experiment.

You will be paid your earnings privately in cash before you leave the lab.

C.10 Robustness Checks

In this appendix, we present the relevant results used for robustness checks for the 8×8 Treatment and the Alt-High Information Treatment.

C.10.1 Aggregate Results

Table C.17: Mistake Rates: Including Timeouts

	O_5	O_8
A_5		
Mean	0.168	0.160
Std Error	0.022	0.021
N	62	62
A_8		
Mean	0.131	0.242
Std Error	0.021	0.022
N	62	62

$p < 0.01$ for $O_5A_8 \to O_8A_8$, $O_8A_5 \to O_8A_8$, $O_5A_5 \to O_8A_8$
$p > 0.10$ otherwise

Table C.18: Mistake Rates: Excluding Timeouts

	O_5	O_8
A_5		
Mean	0.159	0.147
Std Error	0.022	0.021
N	62	62
A_8		
Mean	0.108	0.223
Std Error	0.021	0.021
N	62	62

$p < 0.10$ for $O_8A_5 \to O_5A_8$
$p < 0.05$ for $O_5A_5 \to O_8A_8$ and $O_5A_5 \to O_5A_8$
$p < 0.01$ for $O_8A_5 \to O_8A_8$ and $O_5A_8 \to O_8A_8$
$p > 0.10$ otherwise

Table C.19: Time: No Timeouts

	time	
	O_5	O_{15}
A_5		
Mean	48.935	49.586
Std Error	1.148	1.126
N	62	62
A_{15}		
Mean	51.754	55.345
Std Error	1.276	1.180
N	62	62

p ¡ 0.10 for $O_5A_5 \rightarrow O_5A_8$

p ¡ 0.05 for $O_5A_8 \rightarrow O_8A_8$

p ¡ 0.01 for $O_8A_5 \rightarrow O_8A_8$ and $O_5A_5 \rightarrow O_8A_8$

Table C.20: Time: Timeouts as Maximum Time

	time2	
	O_5	O_{15}
A_5		
Mean	49.124	49.900
Std Error	1.165	1.141
N	62	62
A_{15}		
Mean	52.289	55.784
Std Error	1.266	1.180
N	62	62

p ¡ 0.05 for $O_5A_8 \rightarrow O_8A_8$ and $O_5A_5 \rightarrow O_5A_8$

p ¡ 0.01 for $O_8A_5 \rightarrow O_8A_8$ and $O_5A_5 \rightarrow O_8A_8$

Table C.21: Time: Correct

	timecorrect	
	O_5	O_{15}
A_5		
Mean	48.733	49.209
Std Error	1.096	1.078
N	62	62
A_{15}		
Mean	51.904	54.914
Std Error	1.207	1.279
N	61	62

p ¡ 0.10 for $O_8A_5 \rightarrow O_5A_8$

p ¡ 0.05 for $O_5A_8 \rightarrow O_8A_8$ and $O_5A_5 \rightarrow O_5A_8$

p ¡ 0.01 for $O_8A_5 \rightarrow O_8A_8$ and $O_5A_5 \rightarrow O_8A_8$

Table C.22: Timeouts

	timeout	
	O_5	O_{15}
A_5		
Mean	0.010	0.015
Std Error	0.005	0.005
N	62	62
A_{15}		
Mean	0.024	0.026
Std Error	0.006	0.009
N	62	62

p ¡ 0.10 for $O_5A_5 \rightarrow O_8A_8$

p ¡ 0.05 for $O_5A_5 \rightarrow O_5A_8$

C.10.2 Alt-High Info Results

Table C.23: Willingness to Pay: 15 x 15

	wtp
	$WTP(O_{15}A_{15})$
Mean	5.452
Std Error	0.819
N	31

Excludes one observation where WTP = 70 ECU

Table C.24: WTP Greater Than Zero

	wtp
	$WTP(O_{15}A_{15}$
Mean	0.844
Std Error	0.065
N	32

C.11 Additional Analyses

C.11.1 Additional Aggregate Results

Table C.25: Mistake Rates: Excluding Timeouts

		O_5	O_{15}
A_5	Mean	0.193	0.201
	Std Error	0.013	0.013
	N	222	222
A_{15}	Mean	0.193	0.299
	Std Error	0.012	0.016
	N	222	222

$p = 0.000$ for $O_{15}A_5 \rightarrow O_{15}A_{15}$, $O_5A_{15} \rightarrow O_{15}A_{15}$, and $O_5A_5 \rightarrow O_{15}A_{15}$
$p > 0.100$ otherwise.

Table C.26: Time: Timeouts Treated as Maximum Time

		O_5	O_{15}
A_5	Mean	49.200	50.405
	Std Error	0.713	0.677
	N	222	222
A_{15}	Mean	53.769	57.374
	Std Error	0.779	0.782
	N	222	222

p = 0.00 for $O_5A_5 \rightarrow O_5A_{15}$, $O_{15}A_5 \rightarrow O_{15}A_{15}$,
$O_5A_{15} \rightarrow O_{15}A_{15}$, $O_5A_5 \rightarrow O_{15}A_{15}$, and $O_{15}A_5 \rightarrow O_5A_{15}$
$p > 0.10$ for $O_5A_5 \rightarrow O_{15}A_5$

Table C.27: Time: Correct

		O_5	O_{15}
A_5	Mean	48.240	49.641
	Std Error	0.727	0.662
	N	222	220
A_{15}	Mean	52.615	56.613
	Std Error	0.769	0.776
	N	222	222

p = 0.00 for $O_5A_5 \rightarrow O_5A_{15}$, $O_{15}A_5 \rightarrow O_{15}A_{15}$,
$O_5A_{15} \rightarrow O_{15}A_{15}$, $O_5A_5 \rightarrow O_{15}A_{15}$, and $O_{15}A_5 \rightarrow O_5A_{15}$
$p > 0.10$ for $O_5A_5 \rightarrow O_{15}A_5$
Conditional on Correct

C.11.2 Time Cost Results

Table C.28: Robustness: Time Thresholds

	$t < 73$	$t < 70$	$t < 65$
model			
Options	-0.830	0.433	1.719
	(-0.69)	(0.36)	(1.37)
Attributes	4.078***	3.494***	2.836**
	(3.52)	(2.93)	(2.26)
Options * Attributes	1.392**	0.958	-0.515
	(2.07)	(1.41)	(-0.74)
Period	-0.254***	-0.235***	-0.214***
	(-17.32)	(-15.81)	(-13.97)
Cognitive Score	10.30***	11.06***	12.01***
	(10.80)	(11.38)	(12.11)
Female	-2.605***	-2.429***	-2.484***
	(-7.72)	(-7.11)	(-7.13)
Economics/Business	-2.269***	-2.469***	-2.186***
	(-5.82)	(-6.25)	(-5.46)
English	-3.206***	-2.587***	-2.533***
	(-7.70)	(-6.07)	(-5.79)
Position	0.132**	0.0945*	0.0455
	(2.48)	(1.77)	(0.84)
Positive	-1.270***	-0.955***	-0.871***
	(-3.86)	(-2.88)	(-2.59)
Option Complexity	0.606**	0.259	-0.0289
	(2.05)	(0.86)	(-0.09)
Attribute Complexity	0.116	0.144	0.108
	(0.37)	(0.45)	(0.32)
Constant	53.81***	49.84***	46.05***
	(34.73)	(31.85)	(29.12)
sigma			
Constant	15.06***	14.53***	13.64***
	(127.82)	(121.97)	(112.36)
Observations	8169	7438	6312

t statistics in parentheses

All specifications exclude timeouts

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.29: Robustness: Time Thresholds, Correct

	$t < 73$	$t < 70$	$t < 65$
model			
Options	-0.888	0.377	1.413
	(-0.72)	(0.30)	(1.11)
Attributes	5.049***	4.802***	3.803***
	(4.22)	(3.92)	(2.98)
Options * Attributes	3.119***	2.906***	1.678**
	(4.58)	(4.24)	(2.40)
Period	-0.200***	-0.189***	-0.179***
	(-13.36)	(-12.62)	(-11.73)
Cognitive Score	6.580***	7.024***	8.251***
	(6.79)	(7.20)	(8.38)
Female	-1.263***	-1.061***	-1.244***
	(-3.70)	(-3.11)	(-3.62)
Economics/Business	-2.419***	-2.794***	-2.522***
	(-6.18)	(-7.11)	(-6.43)
English	-2.118***	-1.466***	-1.344***
	(-5.07)	(-3.47)	(-3.14)
Position	0.191***	0.142***	0.121**
	(3.51)	(2.62)	(2.23)
Positive	-0.938***	-0.808**	-0.759**
	(-2.83)	(-2.44)	(-2.29)
Option Complexity	0.529*	0.169	-0.118
	(1.73)	(0.54)	(-0.37)
Attribute Complexity	-0.0952	-0.217	-0.126
	(-0.29)	(-0.65)	(-0.36)
Constant	52.97***	50.46***	46.95***
	(34.09)	(32.48)	(30.27)
sigma			
Constant	13.48***	12.94***	12.01***
	(113.42)	(108.75)	(100.36)
Observations	6432	5913	5036

t statistics in parentheses

All specifications exclude timeouts

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C.11.3 GPA Robustness Checks

Table C.30: Mistake Rate Regressions with GPA

	Model 1	Model 2	Model 3	Model 4	Timeouts as Mistakes
Options	0.00969	0.0127	-0.0202	-0.0532*	-0.0740**
	(0.84)	(1.07)	(-1.49)	(-1.84)	(-2.48)
Attributes	0.000268	0.00364	-0.00355	-0.0143	0.0127
	(0.02)	(0.29)	(-0.28)	(-0.49)	(0.44)
Options * Attributes	0.0871***	0.0785***	0.0852***	0.0869***	0.0923***
	(4.82)	(4.28)	(4.61)	(4.67)	(4.96)
Period	0.000314	0.000276	0.000288	0.000283	-0.00100**
	(0.71)	(0.61)	(0.63)	(0.62)	(-2.10)
GPA		-0.246***	-0.246***	-0.246***	-0.251***
		(-3.98)	(-3.99)	(-3.99)	(-4.04)
Female		0.0935***	0.0935***	0.0935***	0.0909***
		(4.13)	(4.13)	(4.13)	(3.94)
Economics/Business		0.00577	0.00560	0.00559	0.0228
		(0.22)	(0.22)	(0.22)	(0.83)
English		-0.00139	-0.00147	-0.00147	-0.0128
		(-0.06)	(-0.06)	(-0.06)	(-0.51)
Position			0.00452***	0.00499***	0.00598***
			(3.62)	(3.89)	(4.43)
Positive			-0.0304***	-0.0325***	-0.0314***
			(-3.66)	(-3.84)	(-3.63)
Option Complexity				0.00904	0.0134*
				(1.26)	(1.83)
Attribute Complexity				0.00300	0.000322
				(0.39)	(0.04)
Observations	8555	8121	8121	8121	8440

t statistics in parentheses

Marginal effects from logit regression specifications

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.31: Time Regressions with GPA

	Model 1	Model 2	Model 3	Model 4	Correct	Timeouts as Time
Options	2.255***	2.211***	1.230***	-1.270	-1.284	-2.024**
	(6.38)	(6.01)	(2.67)	(-1.38)	(-1.31)	(-2.05)
Attributes	5.108***	5.188***	4.919***	4.133***	5.066***	4.988***
	(12.00)	(11.85)	(11.10)	(4.40)	(4.52)	(5.26)
Options * Attributes	1.483***	1.473***	1.720***	1.830***	3.297***	2.109***
	(3.01)	(2.88)	(3.37)	(3.56)	(5.95)	(4.05)
Period	-0.263***	-0.256***	-0.255***	-0.256***	-0.200***	-0.294***
	(-10.26)	(-9.61)	(-9.60)	(-9.62)	(-9.79)	(-10.05)
GPA		8.896**	8.888**	8.886**	5.341	8.071*
		(2.11)	(2.11)	(2.10)	(1.56)	(1.84)
Female		-2.652*	-2.655*	-2.655*	-1.302	-2.541*
		(-1.89)	(-1.89)	(-1.89)	(-1.13)	(-1.80)
Economics/Business		-1.702	-1.706	-1.708	-1.961	-0.946
		(-1.09)	(-1.10)	(-1.10)	(-1.42)	(-0.60)
English		-3.369**	-3.372**	-3.373**	-2.311*	-3.676**
		(-2.25)	(-2.25)	(-2.25)	(-1.68)	(-2.46)
Position			0.130***	0.162***	0.204***	0.197***
			(3.19)	(3.84)	(4.45)	(4.61)
Positive			-1.221***	-1.380***	-0.957***	-1.355***
			(-4.60)	(-5.11)	(-3.40)	(-4.81)
Option Complexity				0.694***	0.591**	0.840***
				(3.23)	(2.48)	(3.65)
Attribute Complexity				0.220	0.00633	0.128
				(0.94)	(0.02)	(0.55)
Observations	8555	8121	8121	8121	6332	8440

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.32: WTP Regressions with GPA

	WTP			WTP > 0		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Mistakes	0.198**	0.219**	0.319***	0.205*	0.212**	0.437***
	(0.0882)	(0.0864)	(0.111)	(0.108)	(0.0914)	(0.146)
Time	-0.00187	-0.00165	0.000841	-0.00177	-0.00145	0.00180
	(0.00180)	(0.00192)	(0.00287)	(0.00187)	(0.00204)	(0.00256)
Attributes	0.311**	0.311**	0.431	0.0817	0.0727	0.147
	(0.149)	(0.156)	(0.275)	(0.144)	(0.159)	(0.310)
High Info	0.0656	0.150	4.279	-0.167	-0.0936	6.195**
	(0.475)	(0.490)	(2.647)	(0.396)	(0.448)	(3.127)
Female		-0.112	-0.214		0.565	0.397
		(0.439)	(0.447)		(0.414)	(0.419)
GPA		-0.890	-0.682		-1.411	-1.218
		(1.149)	(1.161)		(0.970)	(1.034)
High Info * Mistakes			-0.218			-0.358*
			(0.175)			(0.215)
High Info * Time			-0.00528			-0.00813*
			(0.00401)			(0.00445)
High Info * Attributes			-0.370			-0.347
			(0.331)			(0.340)
Constant	4.485***	4.949***	3.168*	2.539**	3.167**	0.868
	(1.092)	(1.296)	(1.735)	(1.252)	(1.414)	(1.634)
sigma						
Constant	3.225***	3.192***	3.177***			
	(0.210)	(0.212)	(0.209)			
Observations	444	422	422	444	422	422

Standard errors in parentheses

Models 1 - 3: Tobit regression specifications with lower limit of 0 and upper limit of 15

Models 4 - 6: Logit regression specifications

Robust standard errors reported are clustered at the Subject level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C.11.4 Additional Welfare Measures

In order to investigate whether our main result was robust to specifications of welfare loss other than the mistake rate, we additionally conducted aggregate analyses on the rank of the final choice. The variable Rank runs from 1, indicating the worst available option, to 5, indicating the best available option. We replicate our main result, that both unavailable options and irrelevant attributes are necessary to generate a welfare loss.

Because in some decision problems multiple available options (other than the best) had the same payoff, this Rank variable results in some ties. To ensure that our results are not sensitive to these ties, we present below tables with i) ties included, ii) ties rounded to the next highest rank, and iii) observations with ties being dropped. Our main result is robust to all specifications.

Table C.33: Rank with Ties

	O_5	O_{15}
A_5		
Mean	4.687	4.618
Std Error	0.016	0.019
N	2162	2174
A_{15}		
Mean	4.649	4.486
Std Error	0.018	0.021
N	2124	2095

$O_5A_5 \to O_5A_{15}$: $p = .65$
$O_5A_5 \to O_{15}A_5$: $p = .239$
$O_{15}A_5 \to O_{15}A_{15}$: $p = 0$
$O_5A_{15} \to O_{15}A_{15}$: $p = 0$
$O_5A_5 \to O_{15}A_{15}$: $p = 0$
$O_{15}A_5 \to O_5A_{15}$: $p = .452$

Additionally, we can measure the amount of ECU lost as a result of a mistake.

Table C.34: Rank with Ties Rounded Up

	O_5	O_{15}
A_5		
Mean	4.709	4.627
Std Error	0.016	0.019
N	2162	2174
A_{15}		
Mean	4.668	4.510
Std Error	0.017	0.020
N	2124	2095

$O_5A_5 \rightarrow O_5A_{15}$: $p = .554$
$O_5A_5 \rightarrow O_{15}A_5$: $p = .163$
$O_{15}A_5 \rightarrow O_{15}A_{15}$: $p = 0$
$O_5A_{15} \rightarrow O_{15}A_{15}$: $p = 0$
$O_5A_5 \rightarrow O_{15}A_{15}$: $p = 0$
$O_{15}A_5 \rightarrow O_5A_{15}$: $p = .405$

Table C.35: Rank with Ties Dropped

	O_5	O_{15}
A_5		
Mean	4.709	4.627
Std Error	0.016	0.019
N	2162	2174
A_{15}		
Mean	4.668	4.510
Std Error	0.017	0.020
N	2124	2095

$O_5A_5 \rightarrow O_5A_{15}$: $p = .554$
$O_5A_5 \rightarrow O_{15}A_5$: $p = .163$
$O_{15}A_5 \rightarrow O_{15}A_{15}$: $p = 0$
$O_5A_{15} \rightarrow O_{15}A_{15}$: $p = 0$
$O_5A_5 \rightarrow O_{15}A_{15}$: $p = 0$
$O_{15}A_5 \rightarrow O_5A_{15}$: $p = .405$

Because the data generation process results in significant variance across decision problem types in terms of "possible loss" (i.e. distance in ECU of each available option relative to the best option), we first present the average values of potential loss for each option, defined as the Maximum Value in the choice set minus the Value of the available option below applied to the value of the available options themselves (i.e. not to choices of our subjects). From Table C.36, we can see that there are significant differences in potential loss from sub-optimal choice across decision problem types.

Table C.36: Maximum Value - Value

	O_5	O_{15}
A_5		
Mean	5.240	6.900
Std Error	0.606	0.829
N	50	50
A_{15}		
Mean	7.260	6.880
Std Error	0.749	0.847
N	50	50

$O_5A_5 \to O_5A_{15}$: $p = 0$
$O_5A_5 \to O_{15}A_5$: $p = 0$
$O_{15}A_5 \to O_{15}A_{15}$: $p = .177$
$O_5A_{15} \to O_{15}A_{15}$: $p = 0$
$O_5A_5 \to O_{15}A_{15}$: $p = 0$
$O_{15}A_5 \to O_5A_{15}$: $p = 0$

We therefore normalize this loss variable by taking the actual loss for the choice of a subject (i.e. Maximum Value - Value of Choice) and dividing it by the Maximum Value minus the Mean Value of available options in the given decision problem. Again, we replicate our main result, with the exception that this Normalized Loss is somewhat higher in $O_{15}A_5$ than in O_5A_5, though this difference is only marginally significant.

Table C.37: Normalized Loss

	O_5	O_{15}
A_5		
Mean	0.138	0.185
Std Error	0.008	0.010
N	2162	2174
A_{15}		
Mean	0.171	0.231
Std Error	0.009	0.010
N	2124	2095

$O_5A_5 \rightarrow O_5A_{15}$: $p = .334$

$O_5A_5 \rightarrow O_{15}A_5$: $p = .096$

$O_{15}A_5 \rightarrow O_{15}A_{15}$: $p = 0$

$O_5A_{15} \rightarrow O_{15}A_{15}$: $p = 0$

$O_5A_5 \rightarrow O_{15}A_{15}$: $p = 0$

$O_{15}A_5 \rightarrow O_5A_{15}$: $p = .495$

Normalized by dividing by $v_i^* - \hat{v}_i$ for each question i

Bibliography

Gordon M Becker, Morris H DeGroot, and Jacob Marschak. Measuring utility by a single-response sequential method. Behavioral Science, 9(3):226–232, 1964.

Charles Bellemare and Bruce Shearer. Gift giving and worker productivity: Evidence from a firm-level experiment. Games and Economic Behavior, 67(1):233–244, 2009.

Gary E Bolton and Axel Ockenfels. Erc: A theory of equity, reciprocity, and competition. American economic review, 90(1):166–193, 2000.

Pedro Bordalo, Nicola Gennaioli, and Andrei Shleifer. Salience theory of choice under risk. Quarterly Journal of Economics, 127(3):1243–1285, 2012.

Pedro Bordalo, Nicola Gennaioli, and Andrei Shleifer. Salience and consumer choice. Journal of Political Economy, 121(5):803–843, 2013.

Pedro Bordalo, Nicola Gennaioli, and Andrei Shleifer. Competition for attention. The Review of Economic Studies, 83(2):481–513, 2016. doi: 10.1093/restud/rdv048. URL `http://dx.doi.org/10.1093/restud/rdv048`.

Jordi Brandts and Gary Charness. Do labour market conditions affect gift exchange? some experimental evidence*. The Economic Journal, 114(497):684–708, 2004.

J Michelle Brock, Andreas Lange, and Erkut Y Ozbay. Dictating the risk: Experimental evidence on giving in risky environments. The American Economic Review, 103(1):415–437, 2013.

Andrew Caplin and Mark Dean. Search, choice, and revealed preference. Theoretical Economics, 6(1):19–48, 2011.

Andrew Caplin and Mark Dean. Revealed preference, rational inattention, and costly information acquisition. American Economic Review, 105(7):2183–2203, 2015.

Andrew Caplin, Mark Dean, and Daniel Martin. Search and satisficing. American Economic Review, 101(7):2899–2922, 2011.

Jeffrey Carpenter. The sequencing of gift exchange: A field trial. Journal of Economic Behavior & Organization, 139:26–31, 2017.

Matias D Cattaneo, Xinwei Ma, Yusufcan Masatlioglu, and Elchin Suleymanov. A random attention model. arXiv preprint arXiv:1712.03448, 2017.

Center for Food Safety. Victory! USDA releases GMO labeling study. https://www.centerforfoodsafety.org/press-releases/5061/victory-usda-releases-gmo-labeling-study, 2017. URL `https://www.centerforfoodsafety.org/press-releases/5061/victory-usda-releases-gmo-labeling-study`.

Ian Chadd, Emel Filiz-Ozbay, and Erkut Y Ozbay. The relevance of irrelevant information. 2018.

Gary Charness and David I Levine. Intention and stochastic outcomes: An experimental study. The Economic Journal, 117(522):1051–1072, 2007.

Gary Charness, Guillaume R Frechette, and John H Kagel. How robust is laboratory gift exchange? Experimental Economics, 7(2):189–205, 2004.

Gary Charness, Francesco Feri, Miguel A Meléndez-Jiménez, and Matthias Sutter. Experimental games on networks: Underpinnings of behavior and equilibrium selection. Econometrica, 82(5):1615–1670, 2014.

Syngjoo Choi, Edoardo Gallo, and Shachar Kariv. Networks in the laboratory. 2015.

Patricia Cohen, Jacob Cohen, Leona S Aiken, and Stephen G West. The problem of units and the circumstance for POMP. Multivariate Behavioral Research, 34 (3):315–346, 1999.

James C Cox, Daniel Friedman, and Steven Gjerstad. A tractable model of reciprocity and fairness. Games and Economic Behavior, 59(1):17–45, 2007.

Brent J Davis, Rudolf Kerschbamer, and Regine Oexl. Is reciprocity really outcome-based? a second look at gift-exchange with random shocks. Journal of the Economic Science Association, 3(2):149–160, 2017.

Mark Dean and Nathaniel Neligh. Experimental tests of rational inattention. Technical report, 2017a.

Mark Dean and Nathaniel Neligh. Experimental tests of rational inattention. mimeo, 2017b. URL `http://www.columbia.edu/~md3405/Working_Paper_21.pdf`.

Mark Dean, Özgür Kıbrıs, and Yusufcan Masatlioglu. Limited attention and status quo bias. Journal of Economic Theory, 169:93–127, 2017.

Armin Falk. Gift exchange in the field. Econometrica, 75(5):1501–1511, 2007.

Armin Falk and Urs Fischbacher. A theory of reciprocity. Games and economic behavior, 54(2):293–315, 2006.

Armin Falk, Simon Gächter, and Judit Kovács. Intrinsic motivation and extrinsic incentives in a repeated game with incomplete contracts. Journal of Economic Psychology, 20(3):251–284, 1999.

Peter H Farquhar and Anthony R Pratkanis. Decision structuring with phantom alternatives. Management Science, 39(10):1214–1226, 1993.

Ernst Fehr, Georg Kirchsteiger, and Arno Riedl. Does fairness prevent market clearing? an experimental investigation. The Quarterly Journal of Economics, pages 437–459, 1993.

Ernst Fehr, Georg Kirchsteiger, and Arno Riedl. Gift exchange and reciprocity in competitive experimental markets. European Economic Review, 42(1):1–34, 1998.

Emel Filiz-Ozbay, John C Ham, John H Kagel, and Erkut Y Ozbay. The role of cognitive ability and personality traits for men and women in gift exchange outcomes. Experimental Economics, https://doi.org/10.1007/s10683-016-9503-2, 2016.

Urs Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. Experimental Economics, 10(2):171–178, 2007.

Ove Frank and David Strauss. Markov graphs. Journal of the american Statistical association, 81(395):832–842, 1986.

Drew Fudenberg and David K Levine. Fairness, risk preferences and independence: Impossibility theorems. Journal of Economic Behavior & Organization, 81(2): 606–612, 2012.

Xavier Gabaix, David Laibson, Guillermo Moloche, and Stephen Weinberg. Costly information acquisition: Experimental analysis of a boundedly rational model. American Economic Review, 96(4):1043–1068, 2006.

Simon Gächter and Armin Falk. Reputation and reciprocity: Consequences for the labour relation. The Scandinavian Journal of Economics, 104(1):1–26, 2002.

Sen Geng. Decision time, consideration time, and status quo bias. Economic Inquiry, 54(1):433–449, 2016.

Uri Gneezy and John A List. Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. Econometrica, 74(5):1365–1384, 2006.

Charles A Holt and Susan K Laury. Risk aversion and incentive effects. American economic review, 92(5):1644–1655, 2002.

Randall Lewis and Dan Nguyen. Display advertisings competitive spillovers to consumer search. Quantitative Marketing and Economics, 13(2):93–115, 2015.

Juan Sebastian Lleras, Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y Ozbay. When more is less: Limited consideration. Journal of Economic Theory, 170:70–85, 2017.

Kristian López-Vargas. Risk attitudes and fairness: Theory and experiment. working paper, 2015.

R Duncan Luce. Individual choice behavior. 1959.

Paola Manzini and Marco Mariotti. Sequentially rationalizable choice. American Economic Review, 97(5):1824–1839, 2007.

Paola Manzini and Marco Mariotti. Categorize then choose: Boundedly rational choice and welfare. Journal of the European Economic Association, 10(5):1141–1165, 2012.

Paola Manzini and Marco Mariotti. Stochastic choice and consideration sets. Econometrica, 82(3):1153–1176, 2014.

Yusufcan Masatlioglu and Daisuke Nakajima. Choice by iterative search. Theoretical Economics, 8(3):701–728, 2013.

Yusufcan Masatlioglu and Efe A Ok. A canonical model of choice with initial endowments. Review of Economic Studies, 81(2):851–883, 2013.

Yusufcan Masatlioglu and Elchin Suleymanov. Decision making within a product network. Technical report, Working Paper, University of Michigan, 2017.

Yusufcan Masatlioglu, Efe A Ok, et al. Rational choice with status quo bias. Journal of Economic Theory, 121(1):1–29, 2005.

Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y Ozbay. Revealed attention. American Economic Review, 102(5):2183–2205, 2012.

Filip Matejka and Alisdair McKay. Rational inattention to discrete choices: A new foundation for the multinomial logit model. American Economic Review, 105(1): 272–298, 2014.

Nate Neligh. Vying for dominance: An experiment in dynamic network formation. Technical report, Working Paper, 2017.

Jacob L Orquin and Simone Mueller Loose. Attention and choice: A review on eye movements in decision making. Acta psychologica, 144(1):190–206, 2013.

Mark F Owens and John H Kagel. Minimum wage restrictions and employee effort in incomplete labor markets: An experimental investigation. Journal of Economic Behavior & Organization, 73(3):317–326, 2010.

Elena Reutskaja, Rosemarie Nagel, Colin F Camerer, and Antonio Rangel. Search dynamics in consumer choice under time pressure: An eye-tracking study. American Economic Review, 101(2):900–926, 2011.

Jared Rubin and Roman Sheremeta. Principal–agent settings with random shocks. Management Science, 62(4):985–999, 2015.

Navdeep S Sahni. Advertising spillovers: Evidence from online field experiments and implications for returns on advertising. Journal of Marketing Research, 53 (4):459–478, 2016.

Kota Saito. Social preferences under risk: equality of opportunity versus equality of outcome. The American Economic Review, 103(7):3084–3101, 2013.

Yuval Salant and Ariel Rubinstein. (A, f): Choice with frames. Review of Economic Studies, 75(4):1287–1296, 2008.

Bradley T Shapiro. Positive spillovers and free riding in advertising of prescription pharmaceuticals: The case of antidepressants. Journal of Political Economy, 126 (1):381–437, 2018.

Herbert A Simon. A behavioral model of rational choice. The quarterly journal of economics, 69(1):99–118, 1955.

Christopher A Sims. Implications of rational inattention. Journal of Monetary Economics, 50(3):665–690, 2003.

Christopher A Sims. Rational inattention: Beyond the linear-quadratic case. American Economic Review, 96(2):158–163, 2006.

Alireza Soltani, Benedetto De Martino, and Colin Camerer. A range-normalization model of context-dependent choice: A new model and evidence. PLoS Computational Biology, 8(7):e1002607, 2012.

Elchin Suleymanov. Stochastic attention and search. 2018.

Amos Tversky and Daniel Kahneman. Loss aversion in riskless choice: A reference-dependent model. The quarterly journal of economics, 106(4):1039–1061, 1991.