ESSAYS IN APPLIED MICROECONOMIC THEORY

Andrew Copland

A dissertation
submitted to the Faculty of the
Department of Economics in
partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Boston College
Morrissey College of Arts and
Sciences
Graduate School

May 2022

ESSAYS IN APPLIED MICROECONOMIC THEORY

Andrew Copland

Advised by:
Uzi Segal, Ph.D.
M. Utku Ünver, Ph.D.
Lucas Coffman, Ph.D.

Abstract

This collection of papers examines applications of microeconomic theory to practical problems. More specifically, I identify frictions between theoretical results and agent behavior. I seek to resolve these tensions by either proposing mechanisms to more closely capture the theoretical environment of interest or extending the model to more closely approximate the world as individuals perceive it.

In the first chapter, "Compensation without Distortion," I propose a new mechanism for compensating subjects in preference elicitation experiments. The motivation for this tool is the theoretical problem of incentive compatibility in decision experiments. A hallmark of experimental economics is the connection between a subject's payment with their actions or decisions, however previous literature has highlighted shortcomings in this link between compensation and methods currently used to elicit beliefs. Specifically, compensating individuals based on choices they make increases reliability, however these payments can themselves distort subjects' preferences, limiting the resulting data's usefulness.

I reexamine the source of the underlying theoretical tension, and propose using a stochastic termination mechanism called the "random stopping procedure" (RSP). I show that the RSP is theoretically able to structurally avoid preference distortions induced by the current state of the art protocols. Changing the underlying context subjects answer questions–by resolving payoff uncertainty immediately after every decision is made–the assumed impossibility of asking multiple questions

without creating preference distortions is theoretically resolved. To test this prediction, I conduct an experiment explicitly designed to test the accuracy of data gathered by the RSP against the current best practice for measuring subject preferences. Results show that RSP-elicited preferences more closely match a control group's responses than the alternative.

In the second chapter, "School Choice and Class Size Externalities," I revisit the many-to-one matching problem of school choice. I focus on the importance of problem definition, and argue that the "standard" school choice model is insufficiently sensitive to relevant characteristics of student preferences. Motivated by the observation that students care about both the school they attend, and how over- or under-crowded the school is, I extend the problem definition to allow students to report preferences over both schools and cohort sizes.[1] I show that, if students do have preferences over schools and cohort sizes, current mechanisms lose many of their advantageous properties, and are no longer stable, fair, nor non-wasteful. Moreover, I show that current mechanisms no longer necessarily incentivize students to truthfully report their preferences over school orderings.

Motivated by the observation that students care about both the school they attend, and how over- or under-crowded the school is, in "School Choice and Class Size Externalities" I extend the standard school choice problem to incorporate both of these elements. I show that, if students do have preferences over schools and cohort sizes, current mechanisms are no longer stable, fair, nor non-wasteful. In response, I construct an alternative matching mechanism, called the *deferred acceptance with voluntary withdrawals* (DAwVW) mechanism, which improves on the underlying (unobserved) manipulability of "standard" mechanisms. The DAwVW mechanism is deterministic and terminates, more closely satisfies core desirable matching properties, and can yield substantial efficiency gains compared to mechanisms that do not consider class size.

[1]Cohort size is intended as a generalization of school crowding, relative resources, or other similar school characteristics.

In the third chapter, I provide an overview of the history of decision experiments in economics, describe several of the underlying tensions that motivate my other projects, and identify alternative potential solutions that have been proposed by others to these problems. In this project, I add context to the larger field of experimental economics in which my research is situated. In addition to the mechanisms discussed by prior literature reviews, I incorporate and discuss recently developed payment and elicitation methods, and identify these new approaches' advantages and drawbacks.

Contents

List of Tables

List of Figures

ACKNOWLEDGMENTS

This has only been possible thanks to the unwavering support from my advisors. Thank you to Uzi Segal, for your exceptional guidance and words of wisdom, both research related and not. You have been an anchor in the tempest-tossed seas of graduate school, providing crucial feedback through numerous research projects and a safe-haven when the waves became particularly choppy. I cannot thank you enough for your availability in person, through email, over the phone, and via Skype (and the steady stream of math, logic, and observational brain teasers supplied in every medium). Your holistic approach to the discipline has fundamentally shaped how I approach my own learning, teaching, and research. Thank you to Utku Ünver for your unerring recommendations and invaluable advice. Not only have your comments demonstrably improved every project we discussed, your presence uniquely exudes joyfulness. I am indebted to your infectious enthusiasm, endless patience, and instinctive clarity brought to every discussion we have had. Thank you to Lucas Coffman, for being a transformative force in my later graduate years. You not only accepted me as an advisee, but invested more time and effort into my future than I could have imagined. Designing and running a full-scale experiment only occurred thanks to your tireless help. Your dedication to your students' future and success is awe-inspiring. Thank you for your ceaseless efforts, and for your words of advice, comfort, and compassion throughout my research and job market.

In addition to my committee, I would like to thank Tayfun Sönmez, Mehmet Ekmekci, Shakeeb Khan, Hideo Konishi, Bumin Yenmez, and Michael Grubb for both their excellent teaching and their time and advice regarding a number of research projects. Your insights and comments were invaluable not only in improving the quality of the finished products, but also in helping me see problems approaches in ways I would never have considered without their help. I would also like to thank Paul J. Healy, Sam

Stelnicki and the Ohio State University Experimental Economics Lab for insights, guidance, and assistance in conducting the experiment for my job market paper, and to Susanto Basu and the Boston College Economics Department for the research funds to pay subjects. For my development as teacher, I am forever appreciative of the wisdom and guidance provided by Can Erbil and Chandini Sankaran in all aspects of leading a class, and of the insights provided by Paul Cichello, Mark Kazarosian, Matthew Rutledge, and Geoffrey Sanzenbacher, whenever I would appear unannounced at your respective offices with a philosophical or practical teaching question. I would also like to thank Gretchen Rowley, Casey Eaton, Eileen Tishler, and Gail Sullivan for your crucial and much-appreciated support and guidance you have provided, and to Maria DaRosa for your endless generosity.

While at Boston College, I have been immensely fortunate to have been surrounded by such a supportive and welcoming community. I would be remiss not to thank my cohort for all their help throughout the years we've been classmates. The kindnesses you all have extended have immeasurably enriched the past few years and made them thoroughly enjoyable. Thank you to Jean Francois Gauthier and Jason Bowman for your selflessness and patience discussing research (and teaching me Matlab), to Zhuzhu Zhou, Sriram Natarajan, and Kenzo Imamura for the illuminating discussions about all aspects of micro theory, and to Sanghyun Han, Hayley Huffman, and Marco Robles, for your supportive friendship as housemates. Thank you to Marco Brianti, Federico Favaretto, Benjamin Ferri, Laura Gáti, Liyang Hong, Norohiro Komura, Tomohide Mineyama, Alex Poulsen, Lia Yin, and Jordan Bulka for your camaraderie and your invaluable comments throughout so many seminars and classes. I also cannot thank my officemates enough for their guidance and friendship: to Priyanka Sarda, Giri Subramaniam, Navin Kumar, and Sajala Pandey for taking a wide-eyed third-year under their collective wing and helping me navigate the idiosyncrasies of graduate student life in the department, to Tina Letsou and Vera Sharunova for the countless suggestions you have offered that have improved my research, for the conversations about teaching and preparing for presentations, and for your generosity of friendship, uncountably many laughs, delicious baked goods, and unforgettable birthday celebrations, to Valeria Ferraro for your commiseration and assistance navigating the perpetually murky waters of the job market process.

I would not have made it to this point without the dedication of so many amazing teachers and professors I have been lucky enough to learn from. I am forever indebted to the kindness of Carole Hamilton and RJ Pellicciotta who inspired and encouraged my interest in economics, to Rajeev Dehejia, Gavin Finn, Winifred Rothernberg, and the faculty at Tufts University for cultivating my fascination with the subject, and to Lee Craig, who took a chance accepting a blind solicitation from a rising sophomore asking to be a research assistant. Thanks to your generosity, I was formally introduced to the world of economic research. It was your guidance that prepared me for my studies at Boston College.

I also cannot thank Eric Santiago, Firoz Jameel, Seth Johnson, Rebecca Murphy, Marcus Alpert, Samuel Kronish, Michael McKinnon, David Pierce, Eric Reed, and Laura Chanoux enough. Your enduring friendships have been profound influences on every aspect of my life.

Finally, I would like to thank my family, whose support throughout this process has been a source of inspiration, comfort, and support. To Uncle Danny and Aunt Lynnie thank you for your kindness and always offering me a home away from home, to Allison, Christian, Matt, Lauren, and Griffin Brown, thank you for being the siblings I never had, to Aunt Judy, for being a wellspring of support, and to Aunt Debbie and Aunt Carol, for the love, wisdom, and wonderful memories for the past three decades. I am grateful for the early encouragement from Robert Copland and Roger Copland, and thankful to Marie Brown, for among so many other things, welcoming Christian and myself into an otherwise quiet and orderly home, and making the first three weeks of graduate school so memorable and fun. I would like to thank Marilyn Novin for so many beautifully written letters of encour-

agement, Sue and Gillian Zankel for all your support (and delicious babka), and Marc, Lori and Geoff Adelsberg for always welcoming me so graciously to Valhalla, for the engaging discussions about economics, psychology, and philosophy, for your valuable perspective, and for being so understanding of all the stresses of graduate school. It is no exaggeration to say you have supported me from the very beginning of this journey—I still remember submitting the application to Boston College from your computer room.

Thank you to my parents, Marta and Alan, for instilling in me a love of learning. You have steadfastly supported me in every endeavor, something for which I am eternally grateful. I thank you for your values of integrity, empathy and fairness, values that I have sought to reflect in my research, teaching, and life. Thank you for never discouraging me from asking questions, for taking those questions seriously, and for your honest answers and explanations.

And thank you to Sara, for being the most incredible partner and friend possible. Your support has been instrumental every step of the way. Through the many late nights, early mornings, and working weekends, your unwavering encouragement is what kept me going. I am forever grateful for your eternal empathy and understanding whenever unexpected class assignments, paper revisions, or teaching assignment would arise. Your kindness and empathy are an inspiration. Your belief enabled me to continue pressing forward, through the many ups and downs. More than any other cause, this dissertation was written because of your support and love for so many years.

To Sara and my Parents

Chapter 1

Compensation without Distortion: Stochastic Termination and Incentive Compatibility in Preference Elicitation

1.1 Introduction

As a fundamentally empirical science, economics requires high quality data for analysis. Successive generations of economists have iteratively improved on methods for generating this data. Early pioneers surveyed other researchers, asking for their preferences over hypothetical bundles of goods. When concerns were raised over the validity of hypothetical choice data, practitioners responded with incentives: compensating subjects and basing that pay on what choices they made during the interaction. More complex theories required asking more questions, which led to more payments. Fresh concerns highlighted a new problem—multiple payments could lead to the unintentional introduction of portfolio or wealth effects in subject decision making. Once again, experimenters found an answer: pay subjects based on a single decision question chosen at random. However even this is not immune to problems. Instead of making a series of isolated decisions, the single random payment introduces a compound lottery—subjects are effectively facing a lottery over every decision they make during an experiment.

These same theoretical concerns are not limited to experimental economics; many empirical fields utilize agents' subjective beliefs in models. While incorporating these aspects can be useful, some previous papers have raised concerns over the reliability and accuracy of the information as currently derived.[2]

[2]Rooted in earlier works raising concerns over belief elicitation (such as Braden and

Ultimately, this leaves practitioners with a series of difficult methodological trade-offs. One of the more pronounced is what compensation protocol to use when asking subjects a series of questions. Incentivizing thoughtfulness requires subject compensation to be linked directly to the choices they make. Paying for multiple questions can introduce a dynamic dimension to any procedure, since earlier choices endogenously change a subject's income during the experiment, or the portfolio of outcomes they can expect to receive at the experiment's conclusion. Choosing one question at random at the end of the experiment can also distort how subjects view later choices: instead of answering a series of distinct questions, their payment becomes a compound lottery. No matter which mechanism a practitioner ends up choosing, they're forced to make untestable, a priori assumptions about subjects' beliefs and preferences.[3] Problematically, these assumptions can directly conflict with the models or theories the experimenter is seeking to test.[4]

In this paper, I theoretically propose modifying and adopting a version of a payment mechanism used in some game theory experiments to a broader class of decision, social choice, and elicitation contexts. The theoretical upshot is that it creates an environment without the sources of preference distortions mentioned above. In short, instead of waiting until the end to choose the "real" question, this proposed "random stopping procedure" (RSP) sequentially resolves payoff-round uncertainty.[5] More specifically,

Kolstad (1991) and Neill et al. (1994)), recent investigations into methods and the impacts of incentives have come to a variety of conclusions. Examples of these varied (and occasionally contradictory) findings include Hurley and Shogren (2005), Armantier and Treich (2013), Hurley et al. (2007) and Wang (2011); alternatively, see Schotter and Trevino (2014) for a broader review of results.

[3]Many others have pointed out the same concern; the result is so fundamental Azrieli et al. (2018) make a version of it their "Proposition 0."

[4]For example, consider an experiment examining the fit of different non-expected utility theories for choice behavior. Implicitly, analysis requires subjects to make decisions independently (i.e. the choice in one decision is not impacted by the choice in a previous one). At the same time, many non-expected utility theories predict subjects *should* be considering these previous decisions.

[5]In game theory experiments, stochastic termination exploits a folk theorem equivalence result between discounting and probability of continued play. First proposed by Roth and Murnighan (1978), ending the experiment with some probability after every stage-

after each choice is made, the "procedure termination lottery" determines whether that round is played "for real" (in which case the selected lottery is played, payoffs are assigned, and the experiment ends) or the experiment continues. If the termination lottery returns "continue," then the next round's question is asked. Critically, if this happens, the previous decision is never revisited.

There are several theoretical advantages of this design. The termination lottery creates a mechanical justification for payoff-independence between rounds. Portfolio and wealth effects are mitigated since only one question is paid. It also does not create a compound lottery over different rounds: if a subject is answering the n^{th} question, they know all previous $n-1$ choices are no longer "payoff relevant." These common sources of contamination are thus avoided without imposing the same strict assumptions over how subjects evaluate options. In addition, the termination lottery also presents a novel mechanism for more directly altering subject motivation, which can be used to either ensure subjects sufficiently consider the options presented, or to counteract physiological effects such as cognitive fatigue.[6]

At the same time, these theoretical advantages are not without practical limitations. The addition of the termination lottery leads to attrition: many subjects only answer a subset of all possible questions. While careful design can minimize this impact (for example, by front-loading questions most important for testing the primary hypothesis), these costs should not be dismissed.

In order to evaluate performance of the random stopping procedure relative to current payment protocols, I conduct an experiment that elicits preferences over altruistic and other-regarding behavior, a setting previous research has demonstrated suffers from preference distortions. Re-

game repetition simulates an experimenter-imposed discount rate. See Chandrasekhar and Xandri (2017), Fréchette and Yuksel (2017), and Sherstyuk et al. (2013) for more.

[6]This is a condition described in neuropsychology literature where repetitive, cognitively-strenuous tasks can impact the process of decision making. For more, see sections 1.3.3 and 1.8.

sults support the hypothesis that the RSP treatment more closely matches established benchmark giving rates for the same population than the current "best practices" mechanism. Secondary analyses present additional evidence that RSP-elicited preferences are less impacted by distortions induced by previous decisions than the alternative mechanism.

While the "best" elicitation procedure is certainly context-specific, the random stopping procedure offers novel benefits compared to the alternatives. In situations where the extent of preference distortions are unknown or are believed to be significant, the RSP allows experimenters to gather reliable data under much looser behavioral assumptions, adding a useful tool to the practitioner's toolbox.

The paper is organized as follows: section 1.2 describes currently used payment procedures and shortcomings, while section 1.3 describes the proposed mechanism and key design elements. I next compare the theoretical implications (section 1.4) and expected cost and recruitment differences (section 1.5) between the random stopping procedure and currently used alternatives. The repeated dictator game experiment is described and the results are analyzed in sections 1.6 and 1.7 respectively. Section 1.8 explores additional design considerations, and 1.9 concludes.

1.2 Overview of Current Procedures and Problems

One central feature in preference elicitation experiments is the "payment mechanism" which determines how subjects' payments are generated, and as a result, defines the theoretically relevant context surrounding choice questions. Following other papers that examine experimental design structures (such as recent examples Cox et al. (2013), Charness et al. (2016) and Brown and Healy (2018)), these procedures can be grouped into three broad classes:[7]

[7] I use terminology most heavily drawn from Cox et al. (2015). For simplicity, I ignore variations of these procedures (e.g. pay all sequentially after each decision vs. after all

1) One Task (OT): Each subject is asked a single choice question, who is then compensated with their preferred option.

2) Pay All (PA): Subjects answer multiple questions and all choices are paid.

3) Pay One Randomly (POR): Subjects are asked multiple questions, but only one round is randomly selected at the end to be played "for real." (This mechanism is also called the **Random Lottery Incentive Mechanism** (RLIM), or **Random Incentive System** (RIS), in the literature.)

Cox et al. (2015)'s review of experimental methods explicitly describes the incentive compatibility problem in multi-round procedures: "[the one task payment procedure] is nevertheless very interesting because it is the only mechanism that is always (i.e., for all possible preferences) incentive compatible." It is a finding echoed in the seminal work of Azrieli et al. (2018), and similar observations have been made previously; Holt (1986), Karni and Safra (1987), and Segal (1988) all describe these types of distortions. To incentivize truth-telling, compensation must be linked to the choices made by a subject. However, if subjects are asked at least two questions, this leads to an apparent impossibility. If participants are paid based on every decision they make, income and wealth effects can change preferences dynamically throughout the experiment. If only one round is randomly chosen and paid, the experiment creates a two-stage lottery over payoffs, which can itself distort behavior.

The theoretical presence of "cross task contamination," which arises when the existence of earlier tasks (such as previous decision questions) distorts subject behavior in later tasks, requires practitioners to impose restrictive assumptions on the structure of their subjects' preferences. For example, POR experiments must assume subject preferences satisfy *compound*

choices are made) and hybrid methods (e.g. pay a subset of decisions), as none of these variations solve the underlying incentive issues discussed here. See also Azrieli et al. (2018).

independence to avoid distortionary effects of multi-stage lotteries.[8] PA experiments, on the other hand, require the absence of income or portfolio effects in subject preferences.

To add formalization, I introduce two definitions similar to those from Azrieli et al. (2018). Suppose experimental subjects have "general preferences" $\succeq$ (which may depend on external, non-laboratory factors, and cannot be directly observed by an experimenter), and "experimentally induced choice preferences" $\succeq_\omega$. Let ω represent the experimental history up to some round i, incorporating all previously answered decision questions and elicited choices, from rounds $1, \ldots, i-1$.[9]

Definition 1 (Incentive Compatibility (IC))**.** An experiment is **incentive compatible** if, $\forall\ \omega$, for any two options a and b presented in any round i, $a \succeq_\omega b \iff a \succeq b$.

Definition 2 (Cross Task Contamination (CTC))**.** An experiment exhibits **cross task contamination** if there exist two possible histories at round i, ω' and ω'', such that, for two options a and b presented in round i, preferences satisfy $a \succeq_{\omega'} b$ but $b \succeq_{\omega''} a$.

Cross task contamination can arise as long as the experiment's structure allows for these across-round interactions. The practitioner, then, must impose preference assumptions for their design to be incentive compatible; effectively this means assuming that the type of contamination present won't actually lead to different choices by the subject. If these behavioral restrictions aren't actually satisfied by the participants, then deviations between observed preferences $\succeq_\omega$ and subjects' underlying general preferences $\succeq$

[8]Compound independence is defined in Segal (1990). Formally, a two-stage lottery A, which yields simple lottery X with probability α and simple lottery Z with probability $1-\alpha$, is preferred to two stage lottery B, which grants simple lotteries Y and Z with probabilities α and $1-\alpha$ respectively, if and only if X is preferred to Y. For more on this result, see Segal (1992), and more recently, Azrieli et al. (2018). Also related are Cox et al. (2015), Harrison, Martínez-Correa and Swarthout (2015), and Azrieli et al. (2019).

[9]In other words, any deviation between $\succeq$ and $\succeq_\omega$ must be due to effects of the experimental history ω.

threaten the validity of experimental analysis. For completeness, I highlight two critical observations:

Observation 1 (One Task Limitation)**.** Since each subject only provides one data point in one task experiments, all observable variation is generated by choice data between subjects.

Observation 2 (Multi-Round Procedure Limitations)**.** Pay one randomly and pay all methods both require *a-priori* assumptions over subjects' preferences, and as a result, neither are generally incentive compatible.

This paper advocates for the adoption of an alternative experimental procedure that theoretically eliminates far more sources of cross-task contamination than other multi-decision procedures, while still generating intra-subject data. This alternative–called the **random stopping payment procedure** (abbreviated to "random stopping procedure," or RSP)–allows for practitioners to generate incentive compatible elicitation and preference data under far looser assumptions than contemporary methods.

This is achieved by introducing a "termination lottery" (described by equation 1) after every decision is made. Depending on the lottery's outcome, the experiment either ends at that stage–with the most recent choice being played for real and payouts assigned accordingly–or continues to the next decision round. If this second possibility is drawn, then the subject is asked the next decision question; crucially, if this happens, the previous choice is no longer "payoff relevant." That is, if a round isn't selected to be played in its corresponding termination lottery, then that round will not be compensated. Unlike POR methods' round selection mechanisms, the RSP's termination lottery serves to purge previous decisions from relevance before additional preferences are elicited.

1.2.1 Related Literature for Dictator Game Experiment

Experiments that use a repeated dictator game environment with multiple, payoff-relevant decisions are somewhat uncommon. Curiously, many that do use a pay one randomly mechanism without attempting to directly test for the possible existence of contamination across decisions (e.g. White et al. (2019) and Ponti and Rodriguez-Lara (2015)). Even when task independence is considered directly, the tests for contamination generally examine only a subset of decisions, usually without any clear justification for why that subset of decisions was picked.[10] As a consequence, cross task contamination is often assumed away implicitly, despite the previously listed findings of measurable psychological effects of repeated decisions.[11]

Also directly related is the branch of literature that uses dictator games to measure preferences for charitable giving. Within this group is significant variation in payment mechanisms used. Some, such as Benz and Meier (2008) and Ghesla et al. (2019), use a pay all design, mitigating wealth effects by only asking a couple of questions to each subject. At the other end of the spectrum are experiments such as Coffman (2016) and Null (2011) that only pay a subset of all participants (as a way to raise the stakes of decisions).Others dictator game experiments pay a subset of the decisions subjects make, including Eckel and Grossman (1996), Eckel et al. (2018), and Wang and Navarro-Martinez (2019).

Finally, there is a group of experiments that tests the impact of experimental design on dictator behavior.[12] Ghesla et al. (2019)'s nudge treatment, and List (2007) and Korenok et al. (2018)'s "giving" vs. "taking" fram-

[10]For example, Miao and Zhong (2016) validate a repeated dictator game protocol with one-task designs without any payment trade-off. In the one task experiment, dictators choose p to optimally allocate probability between two allocations in a lottery. The two decisions tested are:

OT Question 1: $\{$(\$0 for dictator, \$20 for recipient), p; (\$0 for dictator, \$0 for recipient), $1-p\}$
OT Question 2: $\{$(\$0 for dictator, \$20 for recipient), p; (\$0 for dictator, \$10 for recipient), $1-p\}$

This is despite the fact that four of the eleven decisions made in the repeated setting involve a direct trade-off (where increasing one player's payment reduces the other's).

[11]Note that some multi-decision dictator games do manage to avoid failures of incentive compatibility by examining hypotheses that are more resistant to cross task contamination. (See, for example, Exley and Kessler (2018), and Coffman (2019).) However, even in these instances, contamination could lead to attenuation of the effects of interest, biasing results toward the null result.

[12]Many of these findings were incorporated into the experiment used to test the random stopping procedure.

ing results, highlight the impacts of non-zero donation defaults. Practice questions were included to minimize unintended signaling (see, for example, Andreoni and Bernheim (2009)) and experimenter demand effects (like those described in Cooper and Kagel (2016)'s discussion of Dana et al. (2006*a*)).

1.2.2 Task Contamination in Moral Decisions

Previous experiments have identified various behavioral forces that are likely to be present when participants are asked to make multiple endowment-splitting and/or donation decisions. "Warm glow" giving effects, first defined in Andreoni (1989), describes the idea of "impure altruism" — donations made for the impact on one's own psychological state, and not directly rooted in the donation's effect on the recipient's private utility.[13] "Moral licensing" is a potentially competing factor, which occurs when prior charitable behavior is used to excuse subsequent, more selfish actions.[14] A third, "cognitive dissonance avoidance," describes an individual's desire to make choices that are perceived to be internally consistent in a morally relevant frame.[15]

[13]For examples of experiments investigating the presence of warm glow giving in dictator experiment contexts, see Ottoni-Wilhelm et al. (2017), Liebe and Tutic (2010), and Crumpler and Grossman (2008).

[14]See, for example, Pablo et al. (2013) and Clot et al. (2013) for experiments involving a dictator game, or Meijers et al. (2015) for a more general investigation.

[15]In regards to donations, this can lead subjects to artificially adjust giving such that their final donations reflect the individual's perception of relative worth of the cause compared to others also asked about. Although less directly examined than the two previous psychological biases, see Ploner and Regner (2013), Konow (2010) and Konow (2000) for examples.

1.3 Description of the Random Stopping Procedure

1.3.1 Notation and Preliminaries

Consider an experiment with n total *rounds*. In this paper, I focus attention on cases where each round is a binary choice decision problem, and options are presented as "do you prefer lottery A or lottery B?" Of course, either option could represent an outcome with certainty.[16] Let round i elicit a subject's preferences between A_i and B_i, and denote the corresponding decision problem as $d_i(A_i, B_i)$.

One Behavioral Assumption

I first make one common, but usually only implied, assumption over subject behavior explicit. Assume that individuals' preferences are based on current and (potentially) previous decisions, not beliefs over (unobserved) future choice problems. In round i's decision question $d_i(A_i, B_i)$, preferences between A_i and B_i may be influenced by A_i and B_i directly, and by previous decisions $d_1, \ldots, d_{i-1}$. However, preferences in $d_i(A_i, B_i)$ are not sensitive to subjects' beliefs over future possible questions: preferences in round i are not conditional on beliefs over future rounds.[17]

Assumption 1 (Decision Myopia)**.** For any options A and B presented in round i, beliefs about possible future rounds $i + 1, \ldots, n$ do not impact preferences over A and B. If $A_i \succeq B_i$, this relation is independent of beliefs over future rounds.

Decision myopia is necessary for two major reasons. First, if a subject has sufficiently strong beliefs over future decisions, they might incorporate

[16]This isn't strictly necessary—the random stopping procedure could be used in conjunction with a Holt and Laury (2002)-style price list. The focus on binary choices is to avoid potential confounding incentive effects arising from other elicitation methods (e.g. Zhou and Hey (2018) and He and Hong (2018)) which are not the focus of this paper.

[17]Note, in many settings, this is equivalent to assuming subjects have no information about the future path of the experiment.

these into the early round option evaluation, regardless of the procedure and the accuracy of said beliefs. Second, decision myopia is necessary for preference identification—without it, almost any choice pattern could be explained based solely on how possible future, unobserved outcomes enter the subject's utility. Note that, for these reasons, decision myopia is necessary for any multi-round decision experiment.[18]

1.3.2 Proposed Procedure

This subsection lays out the Random Stopping Procedure method in detail, for any round $i = 1, \ldots, n$.

Round i (i^{th} Question Preference Elicitation)**.** The elicitation question is posed to the subject according to the following method:

> **Step 1** (Present Decision Problem)**.** Begin by presenting decision (d_i) to the subject.
>
> **Step 2** (Record Preferences)**.** The subject indicates their preferred choice, which is recorded.
>
> **Step 3** (Experimental Termination Lottery Draw)**.** Immediately after answering d_i, conduct round i's termination lottery defined by equation 1:
>
> $$L^T \equiv (\text{Play Round i Question}, p_i;\ \text{Continue to Round i+1}, 1 - p_i) \quad (1)$$
>
> If 'Play Round i Question' is selected, then the subject's choice of preferred option in decision d_i is played. Once this occurs, the experiment concludes: no additional decisions are asked, and the subject is compensated. Alternatively, if 'Continue to Round $i + 1$ Question' is selected, then the experiment continues to Round $(i + 1)$ Step 1.

[18]For more on the theoretical importance of decision myopia in POR and PA settings, see Cox et al. (2015).

> Note termination probabilities can be round specific (hence indexing p with i in equation 1), and vary from one round to the next. If the experiment requires one round to be compensated, set $p_n = 1$. If not, set $0 < p_n < 1$.[19]

This termination step removes the most problematic sources of cross-task contamination, restoring incentive compatibility without any behavioral or preference assumptions other than myopia. Intuitively, the reason lies in the sequential resolution of payoff uncertainty: the random stopping procedure mimics a sequence of one task choices that may or may not be compensated directly. Once the 'play or not play' uncertainty is resolved for each given decision, the round is never revisited—imposing a theoretical firewall between each decision.

1.3.3 Behavior, Experimental Compensation, and Choice of $\mathbf{p_i}$

The RSP's primary goal is the creation of an experimental structure that purges potential history-dependent cross task contamination through the termination lottery. Ensuring the termination lottery is properly constructed is therefore crucial. On one hand, setting $p_i > x$ for any $x > 0$ is sufficient to nominally ensure truthful reporting (subjects' expected payoffs are higher when answering truthfully). On the other, low values for p_i might undermine the purpose of linking compensation to decisions. Lower values of p induce, *ceteris paribus*, less motivation per round. A rational subject may decide not to exert effort if $p_i \approx 0$ and the utility difference between the presented options is small or difficult to determine.[20] The "proper" value for p_i is thus inexorably linked to the mechanics of the experiment. In this section, I outline a few general "classes" of termination lottery probability mechanisms.

[19] See section 1.3.3 for further discussion.

[20] Motivational aspects are discussed further in section 1.8.

Option 1 (Constant Conditional Probability (CCP))**.** Hold p fixed for every termination lottery. The probability any given round is played (possibly excluding the final question), conditional on reaching it, is p.

Note that the optimal choice for the conditional probability faced in the final round, p_n, is of particular interest. Specifically, conditional on reaching the final round, it is unclear whether p_n should remain the same as all other rounds, or set equal to 1, ensuring one round is always played. While there are potential advantages to holding $p_i = p_n < 1$, (as this still satisfies incentive compatibility, and reduces expected costs), this risks undermining subject motivation. Guaranteeing one round will always be compensated could improve recruitment, clarity, subject instruction, and data quality.[21]

Option 2 (Constant Ex-Ante Probability (CEAP))**.** Set p such that the ex-ante (before question 1 is asked) probability the experiment terminates after any question is identical ($p = 1/n$). This corresponds to a conditional probability of terminating after question i of $p_i = 1/(n-i+1)$.

From a theoretical perspective, both options induce the necessary separation between stages: in both procedures, once a round isn't played, the elicited choice is no longer payoff relevant. Unless otherwise specified, in all subsequent analysis, I use CCP-style RSP experiments where $p_i = 1/n$ for $i < n$ and $p_n = 1$. This allows for more intuitive comparisons between POR and RSP designs, and $p_n = 1$ avoids the concerns mentioned above. That said, there are likely cases where a CEAP mechanism is advantageous. For example, a CEAP-style termination randomization device (like drawing cards from a deck) might be more intuitively understandable for participants, or the increasing conditional probability might be beneficial in

[21]Furthermore, if subjects aren't necessarily compensated for at least one decision, using an "answer randomly" strategy is no longer strictly dominated by an "answer truthfully" strategy, which may be problematic for testing preferences and certain dominance characteristics.

situations with diminishing intrinsic motivation.

Hybrid Methods

There are, of course, several alternatives to the designs described above. One alternative is to hold p_i fixed for a subset of early rounds before increasing p_i throughout later stages (to combat psychological fatigue). A second is to increase p_i for all $i = 1 \ldots n-1$, but at a slower rate than CEAP.[22] A third is to follow the CEAP design, but set the ex-ante probability below $1/n$, allowing for a higher completion rate than cases where $p = 1/n$.[23]

1.4 Desirable Properties for Elicitation Mechanisms

Individual Rationality:

Individual rationality is satisfied if individuals, once they are able to identify a preferred choice, have an incentive to honestly report their preferences.[24] For a random stopping procedure experiment, individual rationality is satisfied if the probability the experiment ends after any given round is positive ($p_i > 0 \ \forall i$). Note that most payment procedures satisfy individual rationality; failures of this property are usually confined to survey procedures.[25]

[22] Leading to *increasing conditional*, but *decreasing ex-ante* probabilities.

[23] This could be straightforwardly accomplished by defining round i's conditional probability to $p_i = 1/(xn - i + 1)$. Setting $x > 1$ yields ex-ante probabilities $p_i < 1/n$. Note, if x is $e/(e-1)$, the completion rate roughly equals that in a CCP experiment where $p_i = 1/n$.

[24] This is slightly different than the meaning generally used in mechanism design, where individual rationality requires individuals prefer to participate. To align the two definitions, consider an experiment to be individually rational if subjects should prefer to "opt in" by answering questions truthfully over answering randomly or strategically.

[25] Some prior work has found violations of individual rationality caused by elicitation mechanisms. For example, Freeman et al. (2015) and Zhou and Hey (2018) find distortions in price list responses compared to pairwise choice experiments. Karni and Safra (1987), Bohm et al. (2012), and Horowitz (2006) highlight distortions in the Becker-DeGroot-Marschak price elicitation mechanism.

Property 1 (Individual Rationality)**.** If $p_i > 0 \; \forall \; i$, then a subject who knowingly prefers option A to B also prefers to report A as the better option.

Wealth/Income/Portfolio Effects:

Wealth or income effects theoretically exist if the experiment compensates decisions immediately they're made and before additional tasks are completed. Alternatively, if compensation is delayed such that all rounds are paid at the end of the experiment, subject preferences become susceptible to portfolio effects from facing multiple lotteries with varying risks and expected returns. Ultimately, either effect can lead to distortions throughout later rounds. These effects are mitigated either by imposing assumptions over subject preferences, or by basing compensation only on one decision.

Property 2 (Lack of Portfolio or Wealth Effects)**.** The random stopping procedure only compensates subjects based on one decision, and therefore avoids inducing wealth or portfolio effects.

Compound Lottery Effects:

To avoid wealth effect concerns, pay one randomly procedures incorporate a random selection lottery after all rounds are answered. However, this changes the experiment's decision context: subjects are now implicitly tasked with constructing their optimal two-stage lottery from the options presented. While the two situations (choice of a preferred option vs. optimal two-stage lottery) are equivalent for many utility theories, it is not universal. In fact, assuming both reduction and compound independence necessarily imposes the requirement that preferences satisfy the independence axiom.[26] As others have described (Azrieli et al. (2018) and Cox et al.

[26]Specifically, in Segal (1990), compound independence and reduction imply mixture independence. This is applied more explicitly to experimental design problems in Azrieli et al. (2018).

(2015) being two recent examples), POR designs are only incentive compatible if preferences satisfy compound independence.

The motivating theoretical advantage of the random stopping procedure is that it does not require preferences to satisfy compound independence or reduction. Since it cannot be known with certainty ex ante, an RSP experiment is able to avoid the imposition of either axiom. The decision myopia assumption in section 1.3.1 restricts preferences from considering unknown future rounds, while the termination lottery "purges" older rounds from subjects' possible payoff lotteries. By generating and immediately resolving payoff uncertainty, there is no multi-round compound lottery, and thus no divergence between reduction and compound independence. Unlike POR methods, the RSP retains theoretical incentive compatibility for decision makers who satisfy either (or both) axiom(s). This observation leads directly to property 3.

Property 3 (Elimination of Compound Lottery Effect)**.** The random stopping procedure does not generate multi-round payoff lotteries, and therefore does not induce compound lottery effects for decision makers who satisfy either reduction or compound independence.

Intra-Subject Hypothesis Testing

In order to test intra-subject hypotheses, preferences over multiple decisions must be elicited. Specifically, no individual's utility can be estimated beyond a single point without asking multiple questions, described as the "one task limitation" in section 1.2.

The random stopping procedure is unique among payment procedures in satisfying all three properties while also asking subjects multiple questions.

Table 1.1: Properties satisfied by different payment procedures

	Survey	One Task	Pay One Randomly	Pay All	Random Stopping Procedure
Individually Rational		✓	✓	✓	✓
Multi-Round Compatible	✓		✓	✓	✓
No Wealth Effects	✓	✓	✓		✓
No Cumulative Lottery Effect	✓	✓		✓	✓

1.5 Implementation Analysis of RSP, POR, and PA

The largest drawback of using the random stopping procedure is the number of subjects who don't answer every question. In this section, I describe how the termination lottery impacts expected numbers of experimental completions and associated costs.

1.5.1 Proportion of Subjects Completing RSP Experiment

Let N denote the number of subjects who answer at least one question, and $\mathcal{S}$ the number who answer every question ("complete" the experiment). The relationship between N and $\mathcal{S}$ is driven by p_i (the probability of termination immediately after question i) and n (the total number of possible questions).[27] For constant ex-ante probability experiments, where $p_1 = {}^1/_n$, in expectation it would require $n \times N$ recruited subjects for $\mathcal{S}$ completions.

If we assume $p_1^{CCP} = p_1^{CEAP}$, it is straightforward to observe that fewer subjects are needed to yield (in expectation) $\mathcal{S}$ completions in a constant

[27]Of course, $\mathcal{S}$ approaches N as $p_i \to 0$. However, this leads to possible motivation concerns described in sections 1.3.3 and 1.8.

conditional probability design. Specifically, for a CCP-RSP, $\mathcal{S} = N\,(1-p)^{n-1}$. Rearranging gives the necessary initial recruitment:

$$N = \frac{\mathcal{S}}{(1-p)^{(n-1)}} \tag{2}$$

For context, if p is chosen such that $p_i = {}^1\!/_n \forall i < n$, the probability a subject answers every question approaches (as n grows large) ${}^1\!/_e$ from above.

1.5.2 Uncertainty and Subject Recruitment

Of course, a practitioner may be concerned not only with the expected number of completions, but also the variance of completions. For simplicity, note the ex-ante probability any individual makes it to round j is analogous to the probability of drawing $(1-p_i)$ for $i = 1, ..., j-1$. Therefore, define $\rho_j \equiv \prod_{i=1}^{j-1}(1-p_i)$. Using a central limit theorem and corresponding confidence levels, $\mathcal{S}$ is defined by equation (3):[28]

$$\mathcal{S} = \rho_n \times N - z\left(\sqrt{(\rho_n) \times (1-\rho_n) \times N}\right) \tag{3}$$

For more contextualization of this result, see appendix B.

Variation: k-Answers per Question

An experimenter could, alternatively, use an RSP design in place of a one task experiment.[29] Instead of requiring $\mathcal{S}$ total completions, here we require at least k answers for every question without regard for who answers any given question. Further, suppose the pool of N recruited subjects is divided into n subgroups, each with ${}^N\!/_n$ individuals. One subgroup's

[28] Note that ρ is likely to be comfortably between 0 and 1, which enables this approximation. For more, see chapter nine in Grinstead and Snell (1998).

[29] Since the RSP minimizes cross task contamination, it could theoretically be used in place of a one task design at a lower cost.

members begin with question 1, and continue (if not stopped) with questions 2, ..., n. A second group begins with question 2 and continues with questions 3, ..., n, 1. Thus, N/n subjects observe each question first, N/n see each question second, and so on.[30]

Table 1.2 reports the results of a Monte-Carlo simulation showing, for $N = 100$ and $n = 20$, how many times the *least answered* of a simulation's n was reached (effectively representing the minimum k for a simulation). Percentiles correspond with response totals for a simulation's least answered question. In other words, (using the first row from the table), in 90% of simulations, every question was answered at least 54 times.

Table 1.2: Monte Carlo simulation results (500 simulations, 100 initial recruits, 20 questions, constant conditional probability of $p = .05$)

Simulation Results	
Level	**Responses**
10^{th} Percentile	54
5^{th} Percentile	53
1^{st} Percentile	50
Minimum	49

The simulation suggests, if it was necessary to obtain $k = 50$ answers for 20 different questions, an RSP experiment would require roughly $2 \times k = 100$ initial recruits (far smaller than the N necessary for a one task experiment).

1.5.3 Expected Cost Comparisons

Finally, I consider how the incorporation of a termination lottery impacts experimental costs. In general, following the results of equation 2, an RSP experiment should be expected to cost between 2.5 and 3 times that of a POR experiment with the same number of completions. Differences

[30]Note that this represents a sort of "worst-case scenario" without any dynamic adjustment of question order during the experiment. For more, see appendix B.1.

in cost between POR and the equivalent PA or OT experiment vary more widely, as these designs' costs are more heavily impacted by the number of questions asked.[31] For an estimate of the cost of previously published experiments using various methods, and a more detailed discussion of the cost determinants of experimental methods, see appendix B.2.

1.6 Dictator Game Experiment

1.6.1 Experiment Design

To test the effects of different procedures on incentive compatibility and task contamination, it was first necessary to identify a decision context where contamination would exist and be observable. Previous research suggests this occurs in decisions involving social preferences. Specifically, concerns for fairness can lead individuals to condition behavior on previous decisions. While this can be rational (see, for example, the discussion in Machina (1989)), history dependence can lead to violations of incentive compatibility.

A common method of examining social preferences in the laboratory is through the dictator game. The basic structure of the dictator game is straightforward: subjects are paired and only one member of each pair is endowed with a certain amount of currency. This individual is then asked how much of their endowment they would like to "donate" to their partner.[32] There is no strategic interaction between players, so positive transfers are often interpreted as evidence of other-regarding preferences.

There are, however, complicating factors that lead some researchers to argue that positive transfers are not true evidence of benevolence, and are

[31] Adding one more question to a PA experiment increases costs by the expected payment of that question. For an RSP experiment, an additional question has two countervailing effects: on one hand it reduces the likelihood of completion by $(1-p)$, it also decreases p if $p = 1/n$ due n increasing by one.

[32] Subjects are paired anonymously and placed in separate rooms, so individuals are unaware of who their "partner" is.

instead a result of unintended experimental influence.[33] Moreover, observed fragility of giving in some prior experiments to player setting and attention, compounded with an inability to directly control for subject environment (the experiment was conducted online and not in a lab), posed difficulties for implementation.

In response, I modify the "standard" dictator game setting, such that subjects make multiple decisions (i.e. they played the "game" multiple times), and the "recipient" player was always one of a sequence of nonprofit charitable organizations. Not only have prior results suggested subjects are more willing to donate larger amounts to charitable causes than other players (see section 1.7.2 for more detail), this structure also allows for a cleaner test of the impacts of mechanism assignment on elicited preferences.

Participants were divided into three treatments. Two treatment arms (a random stopping procedure and a pay one randomly group) were asked to make up to ten decisions about how much they would like to donate to different charitable organizations. To avoid interactions between behavioral sources of cross-task contamination and order effects, the sequence of charities was held constant for all participants. As a result, different average donation amounts for these two groups is strongly suggestive of differential effects of cross task contamination.

To provide a baseline measure of donation preferences for these organizations, the third treatment group was only given a single donation decision to make, and were unaware of the other nine charities seen by the multi-round groups. Comparing POR and RSP averages against the OT donation averages allows for direct testing of contamination. By definition, the one task procedure is immune to contamination, and thus provides a measure of the population's underlying preferences towards particular organizations. Whichever multi-round mechanism is able to more closely approximate one

[33] Zizzo (2009), for example, cites a variety of dictator game experiments, and argues that giving is driven more by experimenter demand effects than other-regarding preferences.

task donations can thus be interpreted to be less sensitive to preference distortions.

Potential Sources of Cross Task Contamination

There are numerous ways the three effects described in section 1.2.2–warm glow giving, moral licensing, and cognitive dissonance avoidance–might impact behavior in this experiment. If individuals receive psychological satisfaction when a donation is "pledged" (i.e. when a subject selects an amount and continues to the next page), warm glow might lead subjects to increase donations relative to their baseline preferences. Moral licensing could be expected to influence donations in the opposite direction: a large donation to one organization could be used to morally justify smaller donations later, even if the individual feels both organizations are equally "deserving" of donations. The impacts of cognitive dissonance avoidance are more ambiguous, and could influence donations in either direction. Subjects may increase or decrease donations to reflect their perception of the relative worthiness of the organization currently being asked about compared to others. Since prior literature has repeatedly found trust in an organization to be a key determinant of donating, cognitive dissonance avoidance might be expected to increase donations to more reputable, well-known organizations.

This experiment thus presents a scenario with potentially sizable and unpredictable cross task contamination. To be incentive compatible, the net impact of all sources of contamination must be zero. Moreover, since these factors cannot easily be predicted or observed, and the high likelihood of heterogeneous effects of these behavioral factors in a population, there is no clear ex post econometric adjustment that could be used to control for preference distortions in a non-incentive compatible experiment. Instead, to accurately measure subject preferences, cross task contamination must be controlled for mechanically in the experiment's design.

1.6.2 *Implementation*

Subjects were recruited via email from the Ohio State University Experimental Economics Laboratory's (OSU-EEL) pool of registered individuals. After a short description, the recruitment email included a link to the experiment page, which was programmed using the LIONESS software (Giamattei et al. (2020)). A total of 514 individuals provided informed consent, answered all decision questions, and completed the payment survey. It was subsequently noted that some subjects very likely took the survey more than once. All results reported in the next section use a "strict" exclusion criteria for identifying repeaters: any time an IP address is associated with multiple submissions, only the first submission is kept.[34] This left 448 responses: 81 pay one randomly, 152 random stopping procedure, and 215 one task completions.[35]

Subjects in each group were provided instructions that described both the nature of the decision(s) they would be making (which were largely identical for all treatment groups) and information specific to their payment protocol. After completing two practice questions to familiarize them with the display and interface, subjects were presented with their first (or, for those in the one task group, only) decision question. Upon conclusion of all donation decisions, basic demographic information was obtained and subjects directed toward a separate Qualtrics form to enter necessary information for payment.

1.6.3 *Behavioral Predictions*

Although prior literature has provided strong evidence for individual sources of cross task contamination, there is little research that examines

[34] A "weaker" exclusion criteria keeps subsequent submissions if those later submissions used veritably different payment information. All reported results are robust to either criteria.

[35] For more information about how subjects were sorted into treatments, see appendix C.

them together, making direct predictions about the relative direction and magnitude of each type of contamination speculative. As a result, in this paper, I focus on general predictions of where preference distortions should be visible and how they can be identified.[36] The upshot of this approach is that the effects of cross task contamination should be relatively unambiguous. Below, I identify several predicted effects of possible preference distortions:

Differential effects of cross task contamination will lead to different average donations between treatments. Deviations in average donations between treatment is the most direct evidence of preference distortions between groups. The presence of cross task contamination in one treatment would be expected to shift the distribution of donations made to the impacted organization. This observation leads to three, closely related, predictions.

First, the presence of cross task contamination will lead to different average donation amounts between the pay one randomly treatment and random stopping procedure group. This prediction is itself does not take a stance on which more accurately reflects the "true" underlying preferences in the population. Since both groups make multiple decisions, while the POR treatment might be more theoretically exposed to inter-dependencies between rounds, differences between the POR and RSP donations alone is not dispositive. For that, donations must be compared against the one task treatment.

Second, if the random stopping procedure more effectively controls for cross task contamination, average donations in the RSP group should more closely mirror one task donations than POR averages. Due to the randomized treatment assignment, the one task group can be used as a measure of the "underlying" preferences for the subject population.[37] If the RSP treatment mechanically controls for contamination more than the POR mechanism, then subjects in the RSP group should have "experimentally induced preferences" that are closer to their "underlying" preferences.

[36] After all, the relative unpredictability of these behavioral effects is one of the pressing problems with conducting multi-decision donation experiments.

[37] The justification for this is the same as presented in section 1.2.

Third, cross task contamination should impact organizations with stronger and better-known public images more strongly than smaller, less-well known charities. This prediction directly relates to the mechanism behind possible hypothesized contamination.

Once again, it is difficult to definitively predict ex ante what the overall net effect will be positive or negative for each group of organizations. It is possible that large organizations will be hurt by their size, for example, subjects might be more easily impacted by a bystander effect for charities with much larger donor bases. At the same time, there are several possible reasons national organizations' donations will be positively impacted by psychological forces. From an information and trust perspective, the three national organizations are all relatively well known, while the local charities are (at least for most individuals) unknown. Subjects who are highly motivated by trust in organizations should be less sensitive to downward behavioral forces for well-known organizations, buoying donations for national charities with strong reputations. Similarly, if charitable donations are driven, at least partially, by selfish motivations, then we might predict a positive bias towards organizations where the mission and effects are well known. Regardless of the direction of effect, in either case, organization-level characteristics should not only impact a subject's underlying preference to donate, but also how significantly different types of contamination might impact the subject's ultimate decision.

In addition to these theorized effects, cross task contamination should be present in other observable channels of behavior. Offering the option to revise their donation after the "real" charity is determined effectively changes the decision context from a multi-round procedure to a one task experiment. For incentive compatible experimental procedures, this change should not matter. Accepting the opportunity to revise one's donation is evidence of induced distortions caused by the presence of other, potentially relevant, decisions. *If subjects in the pay one randomly treatment are impacted more strongly by cross task contamination than subjects in the random stopping* *procedure treatment, POR subjects revisit their donations and change the donated amount more frequently than RSP subjects.*

1.7 Experiment Results

1.7.1 Primary Results

Figure 1.1: Difference in average donations between pay one randomly and random stopping procedure treatments

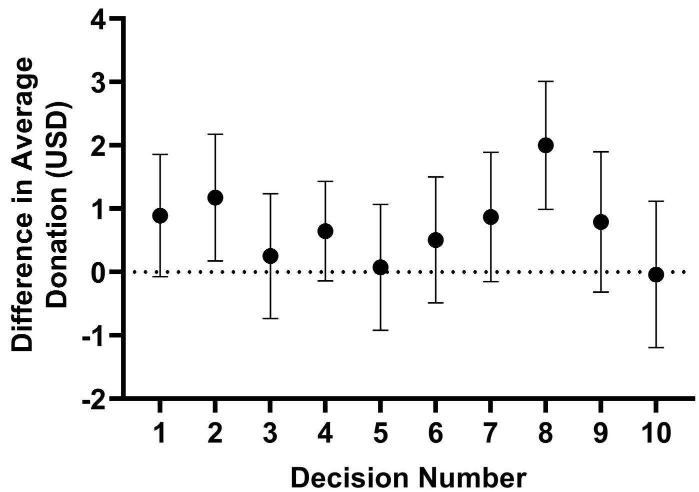

Figure notes: Amount shown is (average POR)-(average RSP) donation. Error bars correspond to 95% confidence intervals.

The first key result is displayed in figure 1.1, which shows the difference between average pay one randomly and random stopping procedure donations for all ten decisions. Statistical analysis using non-parametric tests yields stronger evidence of the significance between POR and RSP donations. As shown in table 1.3 five of the ten donation averages are statistically significantly different at the 10% level using a two-sided Wilcoxon rank-sum test. (Two are significant to 10%, two are significant at 5%, while one is significant $<$1%.)

Result 1: Average donations differ significantly between pay one randomly and random stopping procedure treatments.

Result 1 offers initial evidence of the existence of cross task contamination, and highlights the clear importance behind the choice of payment procedure. Significant differences in average donations appear across a number of organization characteristics.[38] In addition, relative placement in the list order did not appear to have a large effect: POR subjects give more on average to organizations placed both early and late in the list ordering.

Ultimately, result 1 demonstrates the existence of behavioral spillovers in this context. If these spillovers were not present, then both procedures should yield data that were indistinguishable in terms of average donation. This is, however, not what is observed. Significant differences in donation averages should be present only if the two procedures are differentially impacted by cross task contamination. These behavioral deviations between the POR and RSP highlight the importance of establishing which procedure is able to capture underlying preferences more accurately. To do this, I next compare average donations for the random stopping procedure, pay one randomly, and one task treatments for the four organizations included in every treatment arm. Two-sided Wilcoxon rank-sum test and t-test results are given in table 1.4.

Result 2: Random stopping procedure data was unbiased compared to one task decisions, while pay one randomly data yields evidence of significant bias for certain decisions: **Unlike pay one randomly data, which does yield significant deviations compared to one task donation averages for specific decisions, the random stopping procedure data does not significantly differ from any one task donation average.**

[38]More specifically, differences can be observed for both national and local organizations and across multiple focus areas (poverty alleviation and healthcare).

Figure 1.2: Deviations from one task average donations (in USD)

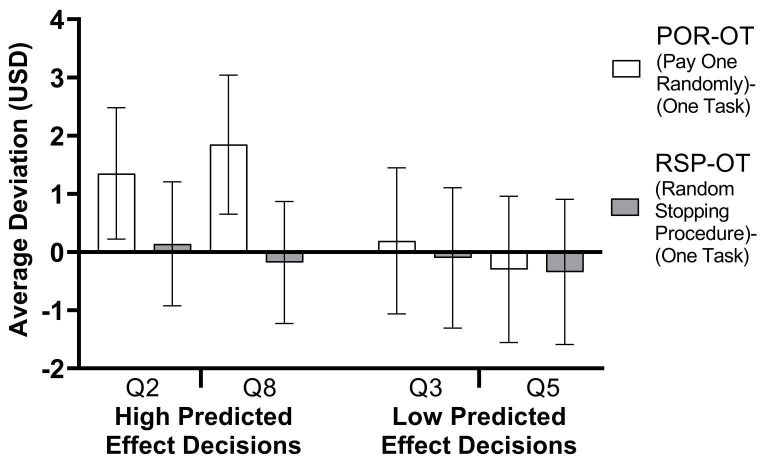

Figure notes: Values correspond to POR-OT and RSP-OT. "High predicted effect" refers to two national-level charitable organizations that are relatively well-known, while "low predicted effect" refers to the two smaller, local organizations. Numbers after each "Q" refers to the decision number from figure 1. 95% confidence intervals are displayed.

Result 2 provides direct support for this paper's underlying theoretical prediction: compared to pay one randomly, data generated by the random stopping procedure more closely matched one task results. Since one task experiments are (by definition) incentive compatible, they can provide useful benchmarks for underlying, population-level preferences. Clear discrepancies between donations is direct evidence of a failure of incentive compatibility per se in the pay one randomly treatment. While the absence of significant differences between one task and random stopping procedure donations is not proof of incentive compatibility, it is strong evidence that the RSP does manage to reduce the effect of cross task contamination observed in the POR treatment.

Moreover, consistent with the third behavioral prediction from the previous section, in both result 1 and 2, the effects of cross task contamination in the pay one randomly treatment appears to have impacted donations differently for national and local organizations.

It is worth noting that these two findings underscore the difficulty associated with predicting the size, direction, and prevalence of cross task contamination without prior knowledge. Observed preference distortions are both widespread and somewhat idiosyncratic. Although some organizational characteristics appear to be more strongly associated with contamination (such as being a national charity), it would be exceedingly difficult to directly control for contamination in a POR experiment without additional information.

I next test the prediction that failure of incentive compatibility should be observable through channels other than average donations. First, I examine the frequencies POR and RSP subjects accepted the offer to revisit, and potentially revise, their initial donation. Results are displayed in figure 1.3 (as well as table 1.5), and provide additional evidence of higher levels of cross task contamination in the POR treatment. More specifically, subjects in the POR treatment were nearly twice as likely to accept the offer to revisit their donation compared to the RSP group (34.6% to 18.0%). We see a similar proportion if only subjects who actually changed their donation are included; 30.9% of subjects in the POR treatment, compared to 16.7% in the RSP group.

> **Result 3: Once the "real" organization was chosen and revealed, pay one randomly subjects were significantly more likely to both revisit and change their previous donation decision compared to random stopping procedure subjects.**

Figure 1.3: Proportion of subjects who accept offer to reconsider donation once real round is revealed for pay one randomly and random stopping procedure treatments

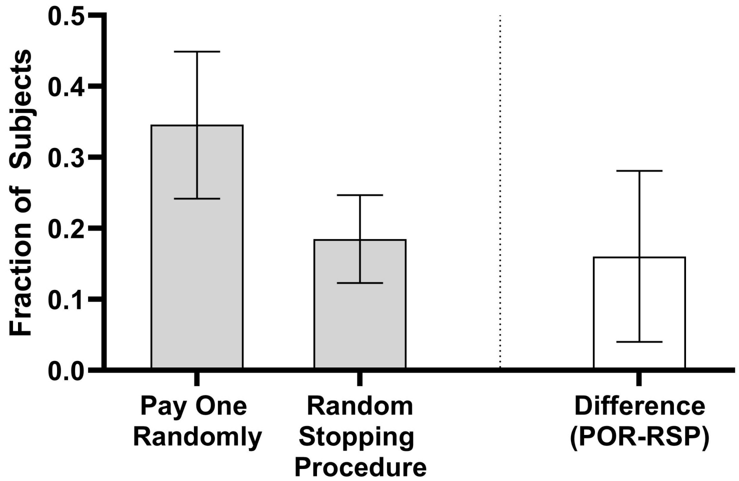

Figure notes: Error bars correspond to 95% confidence intervals.

Result 3 further bolsters the conclusion of result 2: in this setting, cross task contamination induces more significant preference distortions in the POR treatment. As a robustness check for result 3, I excluded RSP subjects who only answer four or fewer questions.[39] These robustness checks yield results consistent with the main finding: the difference in revisitation is not driven by RSP subjects who are only asked a few questions before the experiment ends.

[39] If, for example, as individuals answered more donation questions, they started "experimenting" by selecting suboptimal donation amounts, or grew bored and put less thought into every answer, then the difference in revisitation could be mechanically caused by the fact that many RSP subjects don't answer enough questions to reach the "experimentation" or "boredom" rounds.

Table 1.3: Average giving to each charity for POR and RSP treatments, with both t-statistics and Wilcoxon Rank Sum z-statistics.

	SPHC	DWB	Comm. Health Free Clinic	PETA	Our Companions	Valley Outreach	Community Coalition	Feeding America	AWI	GRASP
				Panel A: Pay One Randomly Average Donation Results						
Donation	5.08	5.18	3.99	2.27	3.99	3.90	4.23	5.20	3.84	3.57
(s.d.)	(3.40)	(3.39)	(3.39)	(2.91)	(3.26)	(3.22)	(3.23)	(3.58)	(3.46)	(3.35)
n	81	81	81	81	81	81	81	81	81	81
				Panel B: Random Stopping Procedure Average Donation Results						
Donation	4.21	4.03	3.76	1.63	3.92	3.39	3.36	3.30	3.05	3.61
(s.d.)	(3.65)	(3.75)	(3.60)	(2.64)	(3.45)	(3.33)	(3.40)	(2.73)	(3.29)	(3.62)
n	152	136	125	113	98	88	83	74	65	61
POR-RSP	**0.87**	**1.15**	**-0.23**	**0.64**	**0.07**	**0.51**	**0.87**	**2.00**	**0.79**	**-0.04**
t-Statistic	**(1.76*)**	**(2.26**)**	**(0.46)**	**(1.59†)**	**(0.15)**	**(1.00)**	**(1.67*)**	**(3.88***)**	**(1.40)**	**(0.06)**
Wilcoxon z-Statistic	**(1.94**)**	**(2.43**)**	**(0.70)**	**(1.76*)**	**(0.29)**	**(1.25)**	**(1.91*)**	**(3.50***)**	**(1.51†)**	**(0.23)**

Notes: † significant at 15%, * significant at the 10% level, ** significant at 5% , *** significant at 1%;
SPHC = St. Petersburg Health Clinic; DWB = Doctors Without Borders; AWI = Animal Welfare Institute

Table 1.4: Differences between average donations by treatment for the four common charities, with both t-statistics and Wilcoxon Rank Sum z-statistics.

	Doctors Without Borders	**Community Health Free Clinic**	**Our Companions**	**Feeding America**
Panel A: Average Donations by Treatment				
POR	5.18	3.99	3.99	5.20
RSP	4.03	3.76	3.92	3.20
OT	3.92	3.81	4.22	3.35
Panel B: Differences Between POR and OT Donations and Statistics				
POR-OT	**1.26**	**0.18**	**-0.23**	**1.85**
t-Statistic	(2.19)**	(0.28)	(0.36)	(3.04)***
Wilcoxon Rank Sum z-score	(2.26)**	(0.41)	(0.21)	(3.09)***
Panel B: Differences Between RSP and OT Donations and Statistics				
RSP-OT	**0.11**	**-0.05**	**-0.30**	**-0.15**
t-Statistic	(0.20)	(0.08)	(0.48)	(0.28)
Wilcoxon Rank Sum z-score	(0.33)	(0.11)	(0.47)	(0.06)
Notes: ** significant at 5% level, *** significant at 1%				

Table 1.5: Decision revisitation measures by treatment group.

	POR (n)	**POR (% of total)**	**RSP (n)**	**RSP (% of total)**	**POR-RSP (pp)**
Panel A: Revisits as a Fraction of Total Treatment Group					
Total Revisits	**28**	**34.6%**	**27**	**18.0%**	**16.6*****
Changed Donation	**25**	**30.9%**	**25**	**16.7%**	**14.2****
Increased Donation	8	9.9%	9	6.0%	3.9
Decreased Donation	17	23.0%	16	10.7%	12.3**
Panel B: Directional Changes in Donations as a Fraction of "Changers"					
% Who Increased Donation		32.0%		36.0%	
% Who Decreased Donation		68.0%		64.0%	
Notes: ** denotes significance at 5% level, *** significant at 1% level; Difference between "total revisits" and "changed donation" caused by some individuals opting to revisit, but not change, the amount they donate.					

Taken together, these results are strongly suggestive of systematic deviations from incentive compatibility for the pay one randomly treatment, but not for the random stopping procedure treatment. Specifically, donation amounts differed significantly between POR and OT treatment in two of the four decisions that could be compared. Moreover, when given the opportunity, more than one third of POR subjects accepted the invitation to revisit their donation once the "real" round was selected. RSP subjects, on the other hand, did not significantly differ from observed OT treatment group behavior. RSP subjects were also more likely to be content with their original donation decision—fewer than one fifth accept the opportunity to revisit their donation once the "real" organization is selected.

1.7.2 Comparison To Previous Experiments

To provide additional context, and better understand the external validity of these estimates, this experiment's average donations and positive-donation frequency were compared to those reported by previous, similar experiments.[40] When asked to split \$10 between themselves and the American Red Cross (in a one-task design), Eckel and Grossman (1996) find decision makers on average donate 31% of the endowment, with just over a quarter of participants donating nothing. Both findings are generally in agreement with this paper's one-task treatment groups who faced either of the two included healthcare-focused charities. Eckel et al. (2018) report that approximately 35% to 39% of the decision makers' endowment is donated, and roughly 87% of their sample gives a positive amount, once again, generally in-line with this paper's experimental results. An average of 45% of individuals' five Euro endowments are donated (and roughly 80% of individuals donate an amount greater than zero) to a predetermined charity in the "house money" dictator game treatment in Wang and Navarro-Martinez (2019). Given that Wang and Navarro-Martinez (2019)'s experiment is a two-round POR-type design (where the other round is a dictator game with another individual as an opponent), their estimates are also broadly aligned with the POR treatment in this experiment, both in terms of donation and

[40]If, for example, subjects in this ten-round charity dictator experiment behaved substantially dissimilarly compared to previous experiments, observed differences between the POR, RSP, and OT treatments might represent idiosyncratic, experiment-specific factors that might not be present in other situations.

proportion of non-zero givers.

1.8 Discussion

Possible Explanations for Observed Patterns in Experiment

There are a number of possible theoretical explanations for the donation patterns observed in this paper's experiment, many of which relate to the factors discussed in section 1.6. Differences between treatment giving behavior could be driven by reference-dependence in donations: once some amount is donated to a smaller organization, it may be harder to psychologically justify giving less to larger, better known charities. For example, if POR subjects are more likely to use early decisions to generate reference points, high giving to the healthcare clinics early on may lead to inflated average donations in subsequent rounds. Alternatively, the POR treatment may be more heavily impacted by a feedback loop of "free" warm glow giving from early decisions. If, for example, early, large donations produced strong positive feelings of psychological wellbeing for POR participants, while the "costs" aren't as salient (since the relevant donation is only realized after every choice is made), it could lead to substantial upward pressure on giving. It is also possible that individuals in the POR and RSP groups simply view the experiment fundamentally differently: POR subjects from the beginning know that one of ten questions will be picked and set donations based on both preferences towards the organization and beliefs about future questions. While under the RSP treatment, even though subjects could perform a similar belief-based donation, the explicit round separation (through the termination lottery observed by subjects after each round) might, at least partially, mitigate this behavior.

It is also worth noting that, while not statistically significant, RSP treatment subjects spent more time, on average, making their decisions.[41] There

[41]More specifically, the average time spent for a POR subject was 180.2 seconds, while

is some research that suggests quicker decisions reflect higher levels of altruism (see, for example Rand (2016) and Carlson et al. (2015)).

It should, of course, be noted that the RSP results do not perfectly align with the one task average donations. There are certainly potential behavioral factors that impact subjects who answer multiple donation questions regardless of the payment procedure used, such as learning about one's own donation preferences, and familiarity or boredom with a repeated, similar task. However, these effects likely exist in the POR treatment as well and are almost certainly unavoidable for any multi-decision procedure. Despite these small deviations, the RSP results more closely aligned with OT behavior in average donations and frequency of positive donation amounts.

Theory and Evidence of Incentive Distortions

Some experiments have suggested that the observed failures of incentive compatibility are actually caused by framing effects. For example, Brown and Healy (2018) suggest that the overall failure of incentive compatibility for a random payment selection method (i.e. a POR experiment) is derived from the common use of price lists, and not from multiple questions creating compound lotteries.[42]

However, there are two reasons for caution when applying this explanation, which may well account for specific observations, more generally. First, there is some evidence that shows systematic manipulation of preferences in pay one randomly experiments even when subjects are presented with pairwise choices (see Cox et al. (2014), for example). Second, the fundamental tension between methodology and theory over the impact of cross task contamination is not resolved. If the problem was "theory predicts subjects should separate tasks, but don't in practice" then the framing explana-

the average time spent by RSP subjects who answered all ten questions was 202.4 seconds. Both averages only include time spent on the ten question screens.

[42] Brown and Healy (2018) and Freeman et al. (2015) find systematic differences in preferences elicited by pairwise choice and by price lists.

tion might fully resolve the conflict. Unfortunately, the issues of CTC are more deeply rooted: many of the non-expected utility models tested suggest individuals shouldn't separate tasks to begin with. Observing that eliminating price lists removes the appearance preference distortions doesn't directly resolve the underlying theoretical issue. Early works like Karni and Safra (1987) and Segal (1988) highlighted that incentive compatibility failed not in spite of various decision theories, but rather these theories predicted a failure of incentive compatibility.[43]

Cognitive Costs and Mental Fatigue

Current experimental literature tends not to explicitly discuss the impacts of costly cognitive processing (cognitive effort costs) and mental fatigue on incentive compatibility.[44] Problematically, decision theories that incorporate costly information acquisition and rational inattention might predict incentive compatibility failures in POR, PA, and RSP experiments.[45]

Unlike current procedures, the RSP design provides an opportunity to partially control effects of fatigue and cognitive effort. For example, suppose a $1/n$ chance of playing any round is unlikely to sufficiently motivate deliberate choice. While a POR experiment provides no alternative, a CCP-style RSP design with $p_i > 1/n$ is both feasible, and provides increased incentive for subjects.[46] Alternatively, a major fear might be subjects becoming bored or tired over time, leading to erroneous choices or disengagement

[43]More recently, it's this theoretical implication that lead Harrison and Swarthout (2014) to suggest, somewhat tongue-in-cheekily, experiments that "selectively enforce" the independence axiom are the work of "bipolar behaviorists."

[44]The neuropsychology literature suggests fatigue might be problematic for experiments that last more than one to two hours (Reteig et al. (2019) and Umemoto et al. (2019)), are cognitively strenuous, or are perceived as boring (Shenhav et al. (2017)). They also document mental fatigue causing cognitive declines (Boksem et al. (2006), Wylie et al. (2017)) and a destabilization effect on economic decision making (Mullette-Gillman et al. (2015)). For theory and biological mechanisms, see Aridan et al. (2019), Massar et al. (2018), Padoa-Schioppa (2011) Peters and Büchel (2009).

[45]If the process of determining one's "true" preferences was taxing, and the expected gains were small, individuals should fall back on an easier, less cognitive resource-intensive heuristic.

[46]This is because $p_i \in (0, 1)$, not $(0, 1/n)$. While setting $p_i > 1/n$ decreases the proportion

from the task at hand. CEAP experiments provide a natural tool for increasing the conditional value of reaching later rounds, offering a means for experimenters to keep subjects engaged and making intentional decisions.

Potential RSP Shortcoming–Certainty Effects

It is worth briefly describing what might happen if the decision myopia assumption is not satisfied. If subjects are significantly motivated by *true* certainty, the random stopping procedure may still cause preference distortions.[47] Choosing A_i over B_i in any round $i < n$ does not mean that the subject will necessarily be guaranteed choice A_i. Instead, it only guarantees that the subject will either receive A_i, or be certain that they will not receive A_i before any other decision is considered.

1.9 Conclusion

By avoiding the between-round separability problems that arise from cross task contamination in all variants of the pay one randomly and pay all designs, the random stopping procedure offers a novel methodology that more closely aligns decision theories of interest with choice-based experimental scenarios. Resolving uncertainty sequentially after each decision through the termination lottery step inhibits the generation of compound lotteries over previous decisions, a necessary feature for incentive compatibility when testing various non-expected utility theories. Once a single decision is compensated, ending the experiment ensures no contamination through an income effect channel. As Azrieli et al. (2018) note, "incentive compatibility is [not] free"—the RSP design is likely to cost, in expectation, two to three times more than an equivalently sized POR experiment to run.

of subjects who answer every question, the value of additional motivation may be worth the cost.

[47]Of course, this problem is not unique to the RSP design: any multi-round decision experiment must either compensate all decisions or face certainty effect concerns.

However, not only is this increase in cost relatively small compared to the POR-PA gap, the random stopping procedure is unique among multi-round elicitation procedures in that it "bears the cost" of incentive compatibility in monetary terms, not by imposing restrictive assumptions over admissible preferences.

Several of these predictions were tested in an experiment that examined POR, RSP, and OT donation behavior in a multi-round dictator game played "against" a sequence of charities. Although the RSP arm required more initial subjects, giving amounts chosen by RSP treatment group more closely matched OT donations than the POR treatment both in terms of overall average donations and the proportion of subjects who donated a positive amount. Additional evidence of greater cross-task contamination in the POR treatment was observed through the POR group's singificantly higher likelihood to choose to revisit their initial donation amount once the "real" round was selected when given the opportunity.

Finally, the RSP offers experimentalists novel tools for controlling more subtle practical concerns. By altering the probability of experimental termination after any given decision, an experimenter can directly account for psychological factors facing subjects, such as fatigue, boredom, or rational inattention.

Chapter 2

School Choice and Class Size Externalities

2.1 Introduction and Motivation

Recent advancements in matching theory have improved the functioning of markets and allocation mechanisms. Extraordinary in their generality and broad applicability, foundational matching algorithms like Gale's Top Trading Cycles (TTC) and Deferred Acceptance (DA) have improved participant outcomes from admissions to exchange markets.[48] Beyond these direct contributions, mechanism design provides a deeper understanding about market function and characteristics across economics as a science (Niederle et al. (2008)).

One underlying result in matching theory is the reality that, while there may exist many "bad" mechanisms, the existence of a singular "best" mechanism depends on the specifics of the problem being studied. Whether it is better to eliminate perceived unfairness or maximize efficiency depends on policymaker preferences. When it comes to school choice problems, numerous works have explored possible partial compromises between the student optimal stable matching (SOSM) and TTC outcomes, allowing for violations of strategy proofness or stability in specific circumstances to improve efficiency (at least if all students are reporting their preferences truthfully).[49] Although these mechanisms have shown substantial promise, all efficiency gains remain constrained by impossibility results.

This paper seeks to highlight that this upper bound on efficiency improvements is itself the result of relaxable structural assumptions imposed onto the problem: coarseness of the measure used to report preferences can

[48] Overviews of these applications can be found in Niederle et al. (2008), Sönmez and Ünver (2009), or Abdulkadiroglu and Sönmez (2013), among many others.

[49] See section 2.2.1 for more detail.

lead to suboptimal allocations. I demonstrate that, by selectively extending the domain of admissible preferences to incorporate additional relevant information, agent welfare can be measurably improved. Specifically, this paper expands on the "traditional" school choice problem by allowing students to define their preference ranking over school/class-size pairs.

There are several reasons this preference extension is proposed. First, I argue that it captures information that is likely relevant for schooling decisions. Second, this extension easily subsumes the standard problem: all "traditional" preference profiles (i.e. only defined over schools) are easily accommodated in this context.[50] Third, depending on student preferences, this extension can yield noticeable utility gains for all students. As described in section 2.2.2, this alternative formulation taps the utility gains received by being matched to a school with empty seats: depending on the distribution of the responsiveness to these gains in the student population, "more class-size-sensitive" students voluntarily relocate to emptier (i.e. underdemanded) schools, potentially freeing up slots at otherwise highly sought-after (overdemanded) schools.

Underpinning this motivation is the documented evidence that students do have preferences over elements of a school like class size, over- and under-crowding, and quality of instructional space. Of particular importance, however, is that the impact of these forces is not uniform across individual students: overcrowded or modular classrooms significantly disrupt some individuals' ability to learn, while others are far less affected. For example, reports have documented the lower air quality often present in modular classrooms, which can disproportionately impact those with allergies or other environmental sensitivities.[51]

If all schools within a district were equally over-capacity, then allowing for preferences over class size would do little good in improving welfare.

[50] Specifically, preferences over school/class-size pairs can be defined lexicographically, as detailed in subsequent sections.

[51] See Board et al. (2004) for more about the health effects of modular classrooms.

However, this tends not to be the case for a variety of reasons. Some school campuses are poorly equipped to handle one method of additional capacity, and therefore relies on another (for example, an inner-city school might not have the physical space for modular classrooms, and would therefore be forced to utilize non-instructional rooms to increase capacity). In geographically larger school districts, population growth can be uneven, and uncorrelated with current school capacity. More recently, demographic forces have led to situations where districts have some schools that are overcapacity, while others are undercapacity, leading to a situation where allowing students to directly consider crowding is particularly relevant.[52] Finally, larger and more crowded schools present differential social factors that may impact students differently.[53]

Finally, as described in more detail in section 2.2.1, it is worth noting that this proposed preference extension remains relevant even if every school is under capacity. As long as preferences are still responsive to the student-body size, student welfare may still be improved by preferential sorting along school and cohort size.

2.2 The Problem in More Detail

2.2.1 Related Literature

Mechanism Design and School Choice

Mechanism design has a long, celebrated history in the mechanics of school choice. Since the earliest works addressing flaws in the contempo-

[52]For example, see "Hillsborough school growth numbers show crowding in some areas, empty seats in others" by Marlene Sokol in the Tampa Bay Times. https://www.tampabay.com/news/education/k12/hillsborough-school-growth-numbers-show-crowding-in-some-areas-empty-seats/2293200

[53]For example, Ready et al. (2004) points to possible adverse impacts of size predicted and explained by the school socialization literature. Lee and Bryk (1989) and Lee et al. (1993) highlight how larger schools may lead to adverse social stratifications and higher levels of student inequality in educational outcomes, a factor that is exacerbated by preexisting socio-economic variation in the student body.

rary allocation methods, matching theory has been applied to different setups of school choice, college admissions, and course allocation problems (see Balinski and Sönmez (1999), Abdulkadiroglu and Sönmez (2003), Sönmez and Ünver (2010), and Budish and Cantillon (2012), among many others). These approaches allow administrators and policy makers to best maximize welfare, while protecting rights and obligations deemed necessary. Administrators principally concerned with avoiding perceived unfairness (by *eliminating justified envy*) might seek to implement a version of a deferred acceptance algorithm. School districts more interested in maximizing efficiency might be better served by turning to a trading cycles-style mechanism (see Pycia and Ünver (2011) and Dur (2012)).

Efficiency Improving Mechanisms

Despite the well-known results of Gale and Shapley (1962), Roth (1982), and Balinski and Sönmez (1999) (among others, see Sönmez and Ünver (2009) for a more complete list), many theoretical works have proposed mechanisms that improve efficiency while minimizing "effective" manipulability. Most notably, Kesten (2010) allows for rejection cycles to be preempted by abridging student preferences under certain circumstances, weakly improving welfare if students are honestly revealing preferences.[54] In addition, empirical analyses have attempted to quantify welfare losses of various mechanisms by using actual preference data from large school districts. For example, Abdulkadiroglu et al. (2009) examines losses from deferred acceptance's use in New York City.

[54]There are many other examples of efficiency-improvements for school choice: Abdulkadiroglu et al. (2015) considers allowing students to indicate intensity of preferences for different schools, Chen and Li (2013) examines a course selection problem and suggests the implementation of a draft system, while also expanding the problem to consider tie breaking procedures, a feature previously examined in Erdil and Ergin (2008).

Another branch of research incorporates the possibility of externalities introduced by other students' choices and final allocations. Although initially focused on the context of firm hiring decisions and other general many-to-one matching problems (see Sasaki and Toda (1996), Saglam and Mumcu (2007), Bando (2012), Fisher and Hafalir (2016), and Pycia and Yenmez (2019), among others), these results and insights have been applied to school choice environments more recently.[55] For the purposes of this paper, these works can be grouped by the nature of externalities considered.

The first, and more broad, inclusion of externalities imposes very few assumptions over agent preferences. (This includes both the "matching with coalitions" literature a la Pycia (2012) and much of the broader body of work on externalities, as in Pycia and Yenmez (2019).) This approach permits preferences not only over the object an individual is matched to, but also the identities of others who are also matched to the same object. (In the school choice context, this would mean students were permitted to hold preferences not only over which school they were allocated, but also the entire set of classmates they have.) Unsurprisingly, many of these results are impossibility theorems. To guarantee existence of a matching, additional structure on preferences must be imposed.[56]

More promising results arise when preferences are restricted to observable characteristics (instead of other individuals' identities). Specifically, in school choice, these papers allow student preferences to depend on the final matched object as well as observable characteristics of the other individuals matched to the same object. Using the running example in Leshno (2021), this could correspond with students caring about both the school and the academic quality of their class-cohort peers (measured by SAT scores, GPA,

[55] Also related is the "matching with contracts" work, such as Hatfield and Milgrom (2005) and Aygün and Sönmez (2012), among others.

[56] For example, Huang (2006) shows that deferred acceptance can be coalition-manipulable. Aksoy et al. (2015) applies this directly to a school choice environment, extending coalition manipulation to study efficiency properties of different mechanisms.

or any other observable metric).[57] Finally, peer effects are estimated in Epple et al. (2018) in the context of modeling district-level decisions of reducing capacity in a school district. In certain ways, this paper and Epple et al. (2018) can be considered complimentary approaches to a similar underlying administrative problem. "Slack" (unfilled seats in particular schools) in a school district cannot be removed without also impacting student preferences over the remaining options. While Epple et al. (2018) model the changes in peer-effects that result from schools closing, this paper highlights the structural effects of changing the number of unused seats in each school.[58]

The most closely related paper is Phan et al. (2021), a project developed independently of this one, which also incorporates the level of crowding into student preferences (specifically, Phan et al. (2021) define preferences over school-resource ratio pairs). In many respects, the two projects are complementary analyses of the same underlying observation: although similarly motivated, subtle differences in approach and restrictions imposed on preferences lead both papers to develop distinct mechanisms that satisfy different properties.

Matching with Contracts

This paper is also more generally related to the body of work examining matching problems with contracts (Kelso and Crawford (1982), and Hatfield and Milgrom (2005)). Specifically, in certain regards, the central algorithm of this paper (a version of deferred acceptance) resembles the iterative, endogenous processes found in those papers (as well as Blum et al.

[57]The effects of peer quality on observed student preferences is examined directly in Abdulkadiroğlu et al. (2020)

[58]The differences in permissible preference structure between this paper and those in the peer-effects literature, namely Leshno (2021), are substantial. For example, Leshno (2021)'s "Example 1" is used to show a scenario where that paper's framework over preferences lead to no stable matching. In this paper, however, that same example leads to a clear stable matching outcome.

(1997)). Despite these appearances, however, results differ substantially between the contracts literature and the problem I examine here. Contrary to the matching with contracts approaches, this setting does not incorporate any aggregate demand assumptions, independence of irrelevant contracts, or substitutability assumptions (see Hatfield and Kojima (2008), Hatfield and Kojima (2010), Echenique (2012), Aygün and Sönmez (2012) for analyses of these conditions, and Sönmez and Switzer (2013) for a matching with contracts environment without bilateral substitutes). Instead, the algorithm I define in Section 2.4 allows for students who have previously proposed to a school are able to tentatively withdraw their application should a better option arise due changes in cohort sizes as the algorithm progresses.

Other Related Work

Many more papers have incorporated welfare-relevant implications of student assignments through channels other than cohort-size or resource ratios, including Troyan (2012) who considers ex-post welfare, and Aksoy et al. (2013) which expands the allocation problem to incorporate cardinal utilities. Experimental works, such as Featherstone and Niederle (2016), identify unexpected efficiency costs of different mechanisms when "played" in a laboratory setting, and could provide an interesting basis for experimental investigation incorporating class-size effects.

Also related are works that have successfully generalized and axiomatized different matching mechanisms, providing crucial insight into the operational advantages and shortcomings of these approaches (see, for example, Kesten and Ünver (2015), Hashimoto et al. (2014), Kojima and Ünver (2014), Dur (2012)). Pathak (2017) surveys theoretical advancements in the field and examines the practical significance and costs associated with satisfying different properties.

Other papers consider alternative aspects of welfare beyond the standard school-choice problem model, including Troyan (2012) who considers ex-post welfare, and Aksoy et al. (2013) which expands the allocation problem to incorporate cardinal utilities. Experimental works, such as Featherstone and Niederle (2016), identify unexpected efficiency costs of different mechanisms when "played" in a laboratory setting, and could provide an interesting basis for experimental investigation incorporating class-size effects.

Finally, this paper is "spiritually similar" to applications where elements of slot differentiation are considered explicitly. For example, Dur et al. (2016) show the importance of considering the order "reserved" slots are filled compared to "open" ones. Similarly, work in cadet-branch matching can be thought of in this light, where certain slots in branches is sought to be "reserved" for different candidates (see Sönmez (2013), Sönmez and Switzer (2013), and Imamura (2021)).

2.2.2 Fluctuations in Over- and Under-crowding of Schools

The observation underlying this paper is twofold. The first is, despite the best efforts of policymakers and school administrators, implementing capacity constraints for schools is inexact and somewhat malleable, should the need for more seats in a district arise. This unfortunate reality has, over the past several decades, led some districts to report substantial rates of *overcrowding*—defined as situations where "the number of students enrolled in the school is larger than the number of students the school is designed to accommodate."[59]

This practical necessity contradicts the commonly-imposed requirement of strict, fixed school capacities. Students newly enrolling in a school district, or moving addresses within a district can put asymmetric stress on any centrally planned allocation outcome. These concerns can be practical (for example, limitations in public transportation can make moving geographically diverse students to a school logistically difficult and time consum-

[59] Definition taken from Lewis et al. (2000).

ing), legal (for example, some school districts set quotas of students in every school that can be eligible for subsidized meals to minimize socioeconomic segregation between schools), or other issues faced by school districts.

The confluence of these factors, combined with shifting population demographics both within and across school districts, has led local governing boards to turn to a variety of possible solutions. Unfortunately for students in these districts, every option presents substantial drawbacks. From utilizing rooms not designed for instruction for that purpose, to moving students to "modular classrooms," to moving schools onto a "year-round" calendar, unanticipated and potentially substantial student welfare losses can be the result.

2.3 Preliminaries And Property Definitions

2.3.1 Notation

A **school choice problem with class size externalizes** is defined over the following elements:

- A set of **students**, S, denoted $S = \{i,\ j,\ ...,\ z\}$
- A set of **schools**, C, denoted $C = \{b,\ c,\ ...,\ h\}$
- A set of **capacities**, Q, which dictates the maximum admissible class size for each school
- A school's **class size** is defined as the number of students holding seats at the school. Note that class sizes are thus a result of an outcome or matching. Denote some school b with x assigned students by $b(x)$. If b's class grew by two students, it would be written $b(x+2)$.
- A set of strict **preferences**, $\succ$, which describes individual student preferences over the set of school-class size pairs. For example, is a student

preferred being the only student in school a to being with three other students in school b, the corresponding notation would be: $a(1) \succ b(4)$.

- A set of strict **priority orderings**, $\blacktriangleright$, which rank the relative admissibility of each student for each school. If student i held a higher priority at some school than student j, it would be denoted $i \blacktriangleright j$

Define a **matching** as a function that maps students onto schools (or, to an outside option $\emptyset$): $\mu : S \rightarrow C \cup \emptyset$, such that, for every school b, the number of students assigned is weakly less then b's maximum capacity.

2.3.2 Two Preference Restrictions

In this section, I introduce two useful assumptions over student preferences. Note that only the first is strictly necessary for the central algorithm of this paper and its associated results.

2.3.2.1 Monotonicity in Class Size

The only strictly necessary restriction imposed on preferences is monotonicity in class size. Specifically, assume that if a school's class size increases, all students view that school less favorably. Formally defined,

Assumption 2 (Monotonicity Assumption)**.** For any school b, preferences for any student i satisfies $b(x) \succ_i b(x+1)$.

This doesn't impose any structure on how different schools may enter the preference ranking for a student. For example, for any two schools b and c,

$$c(1) \succ c(2) \succ ... \succ c(q) \succ b(1)...$$
$$c(1) \succ c(2) \succ b(1) \succ b(2) \succ b(3) \succ c(3)...$$
$$c(1) \succ b(1) \succ b(2) \succ ... \succ b(q) \succ c(2)...$$

all satisfy monotonicity. It is also admissible for a student to only find a partially filled school acceptable (i.e. for a student to have preferences $b(x) \succ \emptyset$ for $x < q_b$). Examples of preferences that violate assumption 2 include:

$$c(2) \succ c(1) \succ c(3) \succ b(1)...$$
$$c(3) \succ b(2) \succ b(1) \succ c(2) \succ c(1)$$

2.3.2.2 *Consistency with the Standard Problem*

In order to compare the school choice with class size externalities to the traditional model of school choice without class sizes, I make one additional assumption. This is not necessary for the algorithm proposed in section 2.4, but is useful for comparing student outcomes induced by the novel algorithm of this paper to alternatives that do not incorporate class sizes.

Assumption 3 (Consistency Assumption)**.** For any student i and two schools b with maximum capacity q_b^{max}, and c with maximum capacity q_c^{max}, $b(q_b^{max}) \succ_i c(q_c^{max})$ in the adjusted framework if and only if the student would have listed $b \succ_i c$ in the standard problem without class sizes.

Note that this assumption only considers preferences over "full" schools.: if a student would have ranked b above c in an environment where the mechanism does not consider class sizes, then the student should consider b to be preferable to c if both schools are full.

2.3.3 *Induced Standard Problem*

Combining the problem definition in Section 2.3.1 with Assumptions 2 and 3 induce what will be denoted a **standard school choice problem** (which I often abbreviated to the "standard problem"). This "standard" framework does not permit students to indicate preferences over class sizes, but is otherwise identical in structure to the one introduced in Section 2.3.1.

For simplicity, I refer to the standard school choice problem as the **standard problem**, and to the school choice problem with class size externalities as the **extended problem**.

Similarly, the deferred acceptance algorithm that does not incorporate class size preferences will be referred to as the "standard deferred acceptance algorithm" or "standard deferred acceptance." The standard deferred acceptance algorithm (defined in appendix 6.B), and does not permit students to withdraw a tentatively accepted offer—the only way a student who is tentatively matched to a school can leave is if another student ranked higher on the school's priority order takes the initial student's place.

2.3.4 Implications of Class-Size Preferences for Traditional Results

Expectedly, the traditional deferred acceptance algorithm loses several of its more useful properties if students have preferences over both school and class size. This is most clearly seen through a violation of non-wastefulness, which is formally demonstrated below:

Proposition 1. If students have preferences over both schools and class sizes, the standard deferred acceptance algorithm violates non-wastefulness.

Proof. Proposition 1 can be proven straightforwardly by example. Consider a scenario with two students i, j and two schools b, c with $q_b = q_c = 2$ (and both i, j are acceptable to both schools). Suppose student preferences are identical:

$$\succ_{i,\, j}\colon b(1) \succ c(1) \succ b(2) \succ c(2)$$

By assumption 3, both students report $b \succ c$ in a "standard deferred acceptance" mechanism. Since both i and j are acceptable to a, both are granted admission and the algorithm terminates. However, both students would prefer $c(1)$ to their final assignment. □

From Balinski and Sönmez (1999), a school choice matching is stable if it satisfies non-wastefulness, individual rationality, and eliminates justified envy. Since the traditional deferred acceptance algorithm is no longer non-wasteful if students hold these extended preferences, Comment 1 directly follows.

Comment 1. If students hold extended preferences, the standard deferred acceptance algorithm's outcome is not necessarily stable.

2.3.4.1 *Justified Envy*

One common notion of fairness in school choice is the *elimination of justified envy* (EJE) property, introduced in Balinski and Sönmez (1999), which is satisfied if, any time a student i prefers another student j's assignment to their own, it must be the case that j has a higher priority than i for that seat. While this property has a good deal of intuitive appeal in the standard framework, it becomes more problematic in the extended problem. Example 1 highlights several of these issues:

Example 1. Suppose there are three students, $i,\ j,\ k$ and two schools, $b,\ c$ with capacities $q_b = q_c = 2$. Student preferences are:

$$i : c(1) \succ c(2) \succ b(1) \succ b(2)$$
$$j : c(1) \succ b(1) \succ c(2) \succ b(2)$$
$$k : b(1) \succ b(2) \succ c(1) \succ c(2)$$

And school priorities are:

$$b : j \blacktriangleright k \blacktriangleright i$$
$$c : j \blacktriangleright i \blacktriangleright k$$

Note there are a total of six possible matchings, which can be categorized by the placement of the "solitary" student.[60] Since j has the highest priority

[60]There are three matchings where a different solitary student is alone in school b and three where they are alone at c.

at both b and c, and j prefers being the solitary student to sharing a school regardless of the two schools in question, the only candidates for a matching that eliminates justified envy is either when j is alone at b or when j is alone at c. Further note that the lexicographic preferences of i and k are such that, for both of these cases, the empty seat would like to be occupied by either i (if j is alone at c) or k (if j is alone at b) Since i and k are acceptable to both schools, any outcome where j is alone is blocked by either (i, c) or (k, b). ■

One direct result of example 1 is summarized in claim 1.

Claim 1. If students have preferences over class size, there is no matching that satisfies both the traditional definition of elimination of justified envy and non-wastefulness.

If we restrict our attention to the "standard" property definitions, claim 1 leads to the following unsurprising conclusion.

Proposition 2. There exists no stable mechanism in the extended-preference school choice problem when using traditional definitions of justified envy and non-wastefulness.

2.3.5 *Extending EJE and Stability*

Incompatibilities between property definitions and modeling extensions are not a new phenomenon—other works have also had to reconsider the appropriate property definitions given changes in the underlying model (for example, Kesten and Ünver (2015) similarly reconsider stability in a non-deterministic school choice problem, while others– such as Ehlers and Morrill (2017)'s notion of *legality*–introduce alternative properties that largely serve the same purpose as stability). Due to the nature of preferences over class size, I follow suit and adapt justified envy to better fit this problem's context.

2.3.5.1 Underlying Tension: Property Rights

The underlying tension behind attempting to directly use the standard justified envy concept in this extended environment with preferences over class sizes arises from ambiguity over the appropriate definition of property rights. In the standard framework, eliminating justified envy is closely associated with protecting the property rights of students: a student's "right" to a seat is derived from their placement on that school's priority ordering. If student i is ranked higher on school b's priority list than student j, property rights are violated if i is passed over for a seat at b in favor of j, despite the fact that i would prefer that seat at b over whatever else they were allocated. In other words, the first q_b students on b's priority order have a "right of first refusal" for a seat; the $q_b + 1^{th}$ student has a right to be the next person in line for a seat if one of the first q_b students chooses not to exercise their right to attend b.

However, when considering circumstances where preferences exist over both class sizes and schools, traditional definitions of justified envy can easily lead to problems, highlighted in the following example (similar to example 1):

Example 2. Consider a situation with three students i, j, and k, and two schools (b and c) each with a capacity of two ($q_b = q_c = 2$). Suppose student preferences are:

$$i : b(1) \succ b(2) \succ c(1) \succ c(2)$$
$$j : b(1) \succ c(1) \succ b(2) \succ c(2)$$
$$k : c(1) \succ c(2) \succ b(1) \succ b(2)$$

And both schools have the same priority ordering of students:

$$b : i \blacktriangleright j \blacktriangleright k$$
$$c : i \blacktriangleright j \blacktriangleright k$$

Finally, suppose the ultimate assignment for each student is defined by μ:

$$\mu = \begin{pmatrix} \{i, j\} & k \\ b & c \end{pmatrix}$$

Of particular interest is student j's assignment, who, despite having a higher priority at both schools, receives a worse outcome (from j's perspective) than k. One potential solution would be trading j and k's assignments, yielding matching ν:

$$\nu = \begin{pmatrix} \{i, k\} & j \\ b & c \end{pmatrix}$$

However, ν represents its own, more substantial, infringement on student property rights. In order to satisfy j, it wasn't sufficient to merely be added to c (as j prefers $b(2)$ to $c(2)$ and would rather stay with i at b). Student k had to be removed from c themselves, despite the fact that c had an open seat, k is acceptable to c, and k prefers $c(2)$ to any matching where they end up at b. ■

2.3.5.2 Extended Justified Envy

To account for situations like example 2, I define an extended notion of justified envy that avoids conflicts of this sort. In effect, a student has rights to a seat in a school if they choose to attend (and can gain admission), but they do not have a right to keep a school artificially under-crowded, or stop another student from taking an empty seat. Note that in cases where schools are full, this new property perfectly mirrors the traditional notion of justified envy.

Definition 3 (Extended Elimination of Justified Envy)**.** *A matching μ eliminates justified envy if:*

Case 1. For all schools in μ where the final number of matched students is equal to the number of seats: There exists no students i and $j \neq i$, and schools b (which is at capacity under μ) and $c \neq b$, where c is either at or below maximum capacity, such that:

1. Student i is ranked higher on the priority list for school b than student j
2. Student i prefers school b at maximum capacity q_b^{max} to c with the class size realized under matching μ: $b(q_b^{max}) \succ_i c(q_c^{\mu})$
3. Student j is matched to b under μ while i is matched to c: $\mu(i) = c(q_c^{\mu})$ and $\mu(j) = b(q_b^{max})$

Case 2. For all schools in μ where the final number of matched students is less than the number of seats: There exists no student i, and no schools b and $c \neq b$, where:

1. $q_b^{\mu} < q_b^{max}$ and $q_c^{\mu} \leq q_c^{max}$ (In other words, under μ, b is under capacity, while c is either at or under capacity.)
2. Student i is acceptable for school b
3. Student i is matched to some other school c under μ: $\mu(i) = c(q_c^{\mu})$
4. Student i would prefer to "join" school b (that is, would prefer school b with a class size $q_b^{\mu} + 1$) than stay at school c with it's final class size: $b(q_b^{\mu} + 1) \succ_i c(q_c^{\mu})$

Intuitively, the first case covers schools that are at capacity and, in such circumstances, captures the same properties as the "standard" definition of EJE. The second case extends justified envy to schools that are below capacity and follows the dynamics described above.

2.4 Deferred Acceptance with Voluntary Withdrawals

The central algorithm this paper proposes is the **deferred acceptance with voluntary withdrawals** (DAwVW) algorithm. There are two equivalent ways of defining the algorithm (one introduces students individually similar in approach to McVitie and Wilson (1971), while the other sorts students all at once like the "standard" deferred acceptance algorithm). Below, I define the version that is used in the proofs throughout the remaining sections of the paper. The second version is included in Appendix C.

The last preliminary step before the algorithm is the notion of **feasibility**. In this context, an option (i.e. a school-class size pair) is **feasible** for a student if that student has not yet "struck" the option from their preference ordering. (The process for striking options is defined in the algorithm.) Intuitively, the best "feasible" option is the best option that has not yet been rejected, either by the student (by withdrawing a previously made offer) or the school.

Round 0: Randomly order all students i, and assign them each a number in line (such that the first student in line is denoted i_1, etc.).

Round 1:

Step 1 (Assignment of Round 1 Contemporaneous Class Size)**.** *For all schools a, define the "round 1 class size" to be the number of students provisionally holding a seat at school a. If a school is empty, assign it a class size of* 1.[61] *Denote the contemporaneous class size of school a at the start of round t by* ξ_a^t.

Step 2 (Application Step)**.** *Assign student* i_1 *to their most preferred option. If the student is acceptable to the school, (s)he is offered a provisional seat. If not, the student "strikes" (i.e. removes) that school from their preference list, and applies to their next most preferred option.*

Round 2:

Step 1 (Round 2 Contemporaneous Class Size)**.** *Update contemporaneous class sizes according to the procedure in round 1, step 1.*

[61] At the start of round 1, all schools will be empty (in this application, at least), and will therefore all have class size 1.

Step 2 (Application Step). *Assign student i_2 to their most-preferred remaining acceptable option. If student i_2 applies to a school where student i_1 is currently holding a seat, continue to round 2 step 3. Otherwise, end round 2.*

Step 3 (Withdrawal Step). *Begin with the student ranked lower in the priority ordering for the relevant school (without loss of generality, suppose this is school a). If there exists another school, denoted b, such that, for the lower ranked student, $b(\xi_b^2) \succ a(\xi_a^2+1)$, then that student voluntarily withdraws from school a. If there is no such option b for the lower ranked student, perform this check with the higher ranked student. If there is no option b that fits this criteria for the higher ranked student, end the withdrawal step.*

If one of the students withdrew from $a(\xi_a^2+1)$, that student strikes $a(\xi_a^2)$ from their preference list (indicating it is no longer a feasible outcome).

If either i_1 or i_2 are unmatched after the withdrawal step, continue to round 2 step 4. If not, end round 2.

Step 4 (Application Step 2). *This application step is reached only if there existed some option $b(\xi_b^2) \succ a(\xi_a^2+1)$ for one student (without loss of generality, suppose this is student i_1). i_1 applies to $b(\xi_b^2)$.*

Round t:

Step 1 (Assignment of Round t Contemporaneous Class Size). *For all schools, define the "round t class size" to be the number of students provisionally holding a seat at the corresponding school. If a school is empty, assign it a class size of* 1*. Denote the contemporaneous class size of school a at the start of round t by ξ_a^t.*

For all students, truncate preferences by "striking" any school-class size option $a(y)$, where $y \leq \xi_a^t - 1$.[62]

Step 2 (Student $\mathbf{i_t}$ Proposing). *Student i_t applies to their top, non-struck option, denoted $s(\xi_s^t)$. If i_t is acceptable to school s, $|\xi_s^t| < q_s$, and there are already $\xi_s^t > 0$*

[62]For example, if there are 6 students holding a seat at school a at the start of round t (i.e. if the start of round t class size for school a is 6), all students strike $a(5),\ a(4), \dots,\ a(1)$ from their preference lists (if they have not been struck already).

students holding a seat at school s, continue to the withdrawal step. If $|\xi_s^t| = q_s$, then the student ranked lowest in school s's priority order is removed from school s and $s(q_s)$ is struck from their preferences, and continue to the next application step.

Step 3 (Withdrawal Step)**.** *Begin with the student ranked lowest in the priority ordering for the relevant school s. If there exists another school, denoted b, such that, for the lowest ranked student, $b(\xi_b^t) \succ s(\xi_s^t + 1)$, then that student voluntarily withdraws from school s. If there is no such option b for the lower ranked student, repeat this process with the next lowest ranked student. Repeat this sequentially, moving from the bottom to the top of the school's priority order, until either one student withdraws, or there is no option b that fits this criteria for the any student at s, in which case, end the withdrawal step.*

If one of the students withdrew from $s(\xi_s^t + 1)$, that student strikes $s(\xi_s^t)$ from their preference list (as it is no longer a feasible option).

If any student $i_1 \ldots i_t$ is unmatched after the withdrawal step, continue to the next application step. Else, end round t.

Step 4 (Application Step 2)**.** *This application step is reached only if there is a student who is unmatched after either the initial application step or the withdrawal step. The unmatched student (without loss of generality, suppose this student is i_t) then applies to their next most preferred option, denoted $r(\xi_s^t)$.*

If i_t is acceptable to school r, $|\xi_r^t| < q_r$, and there are already $\xi_r^t > 0$ students holding a seat at school r, continue to the next withdrawal step. If $|\xi_r^t| = q_r$, then the student ranked lowest in school r's priority order is removed from school r and $r(q_r)$ is struck from their preferences, and continue to the next application step.

Step 5 (Withdrawal Step 2)**.** *Perform the same procedure as defined in the previous round t withdrawal step, now considering the students holding a seat in school r.*

If any students remain unmatched after the second withdrawal step, continue alternating application and withdrawal steps until all students

$i_1, \ldots, i_t$ are holding a seat at a school. Once this is satisfied, continue to round $t + 1$.

2.5 Properties of Interest

2.5.1 *Implications of DAwVW Algorithm Result*

First, make note of two observations, which will be used in later proofs:

Observation 3. At the start of round t, every student who has entered in rounds $1, \ldots, t-1$ is being held by their best feasible option.

Observation 3 follows directly from the assumption that round $t-1$ terminated, a necessary condition for round t to begin. If some student was not holding their best available option at the end of $t-1$, they would have withdrawn from their tentative match and applied to the more preferred option before round $t-1$ concluded.

Observation 4. For any school b in round t, if a student application in step s leads to a withdrawal in the same round, it must take place in step $s+1$. Moreover, the size of b's class immediately before s and immediately after the withdrawal in $s+1$ must be the same.

The intuition for observation 4 also directly follows from the definition of the algorithm and observation 3. Since a school only changes size after a new student applies, only those students tentatively holding a seat at that school find themselves in a situation where some other option may yield a higher utility than the school they are currently matched to, plus a class size of one (from the newly applying student). If all individuals prefer the larger class than any other option, there are no withdrawals and the algorithm moves to the next round. If these is an individual who would prefer some alternative option, they immediately withdraw to pursue it.

Given these two observations, I prove several lemmas that are used in proofs of later propositions (including that the algorithm itself terminates). On their own, they provide insight into the dynamics of the algorithm.

Lemma 1. In any step s in any round t, among the first t students, there is no more than one student not being held by a school (or their outside option).

Proof. Consider a procedure at the start of round i. By the definition of the algorithm, all students except student i, must be held by their most preferred feasible option. Suppose student i applies to some school b with σ_b^i students currently being held by b. By construction of the application step, this means all students in the set holding a seat at b plus student i prefer $b(\sigma_b^i)$ to any other feasible option. After student i applies to b, there are $\sigma_b^i + 1$ students total seeking a seat at b. If all students prefer $b(\sigma_b^i + 1)$ to all other feasible options, no student withdraws and the round terminates. If there are some who would prefer another feasible option over $b(\sigma_b^i + 1)$, then the student who has such a more preferred feasible option ranked lowest on the priority order of b withdraws first (denote this withdrawing student as j, who can be anyone in the set $\{\sigma_b^i,\ i\}$).

Note that immediately after j withdraws, there are once again σ_b^i students being held by b. As mentioned previously, all students in that set must prefer $b(\sigma_b^i)$ to any other feasible option. As a result, no additional student withdraws from b in the same step. The same pattern occurs after the next application step: if j applies to a different school c, at most one student will withdraw from c as a result.

For Lemma 1 to be false, it must be the case that, after some application step, two students withdraw. However, this would contradict the previous result: once a student withdraws, the school's class size returns to the pre-application step size, removing the incentive for any student to withdraw. □

Lemma 2. No school will grow by more than one student in a given round.

Proof. The proof of Lemma 2 follows directly as an implication of Lemma 1. For any school b to have grown by more than one student in round t, it must be the case that, over the course of round t, two students applied to b, were both provisionally accepted by b and did not induce any other student to withdraw from b as a result. (Otherwise, if a student did withdraw from the school after this student's application, it would not have grown, but merely shifted the set of students who were provisionally being held be the school.) However, this would contradict the construction of the algorithm: the round ends when no student is provisionally unmatched, which means round t would end as soon as the first of the two students applied to b without any withdrawals. □

The following lemma highlights one of the key differences between the DAwVW algorithm (and the school choice problem with class size externalities) and the broader matching with contracts literature. Once a student has applied to a particular school, subsequent additional classmates can lead that student to prefer an alternative option. This setting explicitly allows the proposing student to rescind their offer in favor of a better alternative.

Lemma 3. In any round, any given student will withdraw from a school no more than once.

Proof. Consider school b at the start of round t with beginning of t class size of $b(\sigma_b^t)$. By definition of the algorithm, no withdrawal from b has yet occurred in round t. Consider every possible case in turn:

Case 1.1. No student applies to b in round t. By assumption 2 and observation 3, students are being held by their most preferred option, and no student withdraws from $b(\sigma_b^t)$.

Case 1.2. A student previously unmatched to b, denoted i, does apply to b in round t. If this occurs, one of several possible outcomes results:

Sub-Case 1.2.1. All students being held by b at the start of t and i all have preferences that satisfy $b(\sigma_b^t + 1) \succ c(\sigma_c^t)$ for any school $c \neq b$. By definition of the algorithm, no student withdraws in the subsequent withdrawal step and the round terminates.

Sub-Case 1.2.2. There exists at least one student in the set of students being held by b ($\{\sigma_b, \cup i\}$) for whom there exists an option c such that $c(\sigma_c^t) \succ b(\sigma_b^t + 1)$. Denote the set of students for whom this condition holds as ρ_b^t. If the set ρ_b^t is a singleton, that student withdraws from b and strikes $b(\sigma_b^t)$ from their preference list. If there are multiple students in ρ_b^t, then the student in ρ_b^t ranked lowest on b's priority order withdraws and strikes $b(\sigma_b^t)$ from their preference list.

Consider some student j who has already withdrawn from $b(\sigma_b^t)$. If this occurred, it must have been because another student applied to b (increasing its contemporaneous class size to $\sigma_b^t + 1$) and there existed some other school $c(\sigma_c^t)$ that j preferred to $b(\sigma_b^t + 1)$. If j never reapplies to b in this round, then Lemma 3 is trivially satisfied.

Suppose at some step s of round t student j does reapply to school b. By construction of the algorithm, when j withdrew from b the first time, $b(\sigma_b^t)$ was struck from j's list of feasible options. Thus, when j reapplies to b it must be as $b(\sigma_b^t + 1)$. By definition of the set ρ_b^t, j is no longer a member: by reapplying to b as $b(\sigma_b^t + 1)$, it must be the case that there are no feasible options $c(\sigma_c^t) \succ b(\sigma_b^t + 1)$ (otherwise j would apply to $c(\sigma_c^t)$). By Lemma 2, b will not grow beyond $\sigma_b^t + 1$. Thus, j will not withdraw from $b(\sigma_b^t + 1)$.

□

Finally, it can be shown that classes weakly increase as the algorithm progresses:

Claim 2. All contemporary class sizes weakly increase as the round number increases.

Proof. Suppose not, and consider school b. Denote the set of students being held by b at the start of round t as σ_b^t. Define the set of students who hold a seat at b at the end of round t as ψ_b^t. Assume that b is a school where the class size shrinks between the start or round t and $t+1$, thus $|\sigma_b^t| > |\psi_b^t|$. Finally, define the student who enters the procedure in round t as i. Consider the following cases that define the possible courses of round t, each of which leads to a contradiction:

Case 2.1. Student i's most preferred, non-struck option is $\emptyset$. Student i then enters the procedure and exists to their outside option. Given that no applications to schools took place, and given observation 3, no school's set of matched students changes and all school remain the same size.

Case 2.2. Student i prefers some school-class size pair $b(|\sigma_b^t|) \succ \emptyset$. Depending on the preferences of students in $\{\sigma_b^t,\ i\}$, one of the following occurs:

Sub-Case 2.2.1. If all students in the set σ_b prefer $b(|\sigma_b^t + 1|) \succ c(|\sigma_c^t|)$, where $b \neq c$, then all students plus i remain at b and no other school's class size changes (while b grows by one student).

Sub-Case 2.2.2. Suppose $c = b$ and at least one student in $\{\sigma_b^t,\ i\}$ has preferences that satisfy $d(\sigma_d^t) \succ b(\sigma_b^t + 1)$. Following round t step 3, the student ranked lowest on b's priority list, for whom such an option $d(\sigma_d^t) \succ b(\sigma_b^t + 1)$ exists, withdraws from b and applies to that more preferred option $d(\sigma_d^t)$. Once this student withdraws and strikes $b(\sigma_b^t)$ from their preference list, class b's size is once again σ_b^t. Moreover, as the student who withdrew from b has not yet applied to $d(\sigma_d^t)$, all schools other than b are still holding the same classes they were at the beginning of round t. By observation 3 and the construction of i's preferences in case 2, all students remaining in b prefer it to any other option.

The student who withdrew from $b(\sigma_b^t)$, denoted student j, then applies to their most preferred remaining option $d(\sigma_d^t)$ in step 4. If all students in set $\{\sigma_d^t \cup j\}$ all prefer $d(\sigma_d^t + 1)$ to any other non-struck option, then j is granted a temporary seat in d, no student withdraws, and the size of d increases by one student between the start and conclusion of round t. If at least one student holding a seat at d would prefer some other option $e(\sigma_e^t)$ to

$d(\sigma_d^t + 1)$, the lowest ranked student with such preferences withdraws from d and applies to e in step 5.

Using observation 3, it must be the case that this chain of applications and withdrawals is finite. At some point, one of the following occurs:

1. The "applying" student applies to a school (denoted z) where they and the set of all students holding a seat at z prefer the enlarged class to all other feasible options. If this occurs, no student withdraws from z, round t terminates with school z having grown by one student and all other school remaining the same size as they were at the start of round t.
2. For some student, there is no better feasible option than remaining unmatched. If this occurs, that student "matches" to their outside option, no withdrawals occur, and the round terminates without any school changing contemporary class size.

□

2.5.2 Algorithm Termination

In this section, I use the previous observations and claims to demonstrate the algorithm terminates with certainty, formalized in Proposition 3.

Proposition 3. *The deferred acceptance with voluntary withdrawals algorithm terminates.*

Proof. Suppose the algorithm does not terminate; this would require an infinite cycle in some round of the algorithm. Denote this non-ending round as round τ. In order for round τ to have begun, round $\tau - 1$ must have previously ended. By Observation 3, all students numbered 1 through $\tau - 1$ are holding their best feasible option. By construction of step 1 in round τ, all students strike infeasible options (removing them from their preference lists). In step 2, student τ applies to their most preferred, feasible option (denote this school b with start-of-round class size ς_b^τ).

In order for τ to continue indefinitely, it must be the case that some student withdraws from the school that was just applied to. By Lemma 1, only one student would withdraw, and by the construction of step 3 in the algorithm, that student who withdraws from $b(\varsigma_b^\tau)$ strikes $b(\varsigma_b^\tau)$ from their preference ordering as not feasible. Similarly, by construction, in every subsequent withdrawal step, the withdrawing student's preference list shrinks by one option. By Lemmas 2 and 3, since no school grows by more than one student in round τ and no student withdraws from the same school more than once in round τ, there is a limit to the number of withdrawal-application steps in round τ. However, this contradicts the assumption that round τ continues indefinitely. □

2.5.3 *Mechanism is Individually Rational*

Though this mechanism's deferred acceptance structure makes satisfying this property relatively straightforward, it is addressed here for completeness. A mechanism is individually rational if all agents' assignments are better than what they would have received had they not participated in the market.[63] Unlike the standard problem, which only considers school assignment, here individual rationality also applies to class size. To satisfy individual rationality in this problem, students must prefer their school-class size pair outcome over their outside option.

Proposition 4 (Individual Rationality)**.** *The DA with Voluntary Withdrawals algorithm is individually rational.*

Proof. Demonstrating DAwVW satisfies individual rationality is relatively straightforward. By Proposition 3, the DAwVW algorithm ends with a

[63]More specifically, all agents prefer their assignments to their outside options.

matching. For the algorithm to terminate, there cannot be any unmatched student at the end of the final round. If a student preferred their outside option to their tentative match at this stage, they would have voluntarily withdrawn and applied to their outside option. More directly, a student preferring their outside option to their final allocation is preempted by the design of the withdrawal step. □

2.5.4 *Non Wastefulness*

Non-wastefulness is an additional property that straightforwardly extends to this context. Specifically, a matching is non-wasteful if there exists no student who both prefers a currently empty seat at school b to their currently assigned seat in school c (where $b \neq c$) and the student is acceptable by school b.

Proposition 5 (Non Wastefulness)**.** *The DA with Voluntary Withdrawals algorithm is non-wasteful.*

Proof. The proof of DAwVW satisfying non-wastefulness is similar to the one for Proposition 4. Assume that there is a "wasted" seat such that there exists a school b that: 1) at the conclusion of the DAwVW algorithm has class size $\sigma_b^T < q_b$, 2) a student i who prefers $b(\sigma_b^T + 1)$ to their final assignment, and 3) i is acceptable to b. By assumption, this would imply i, before the end of the algorithm, had chosen not to withdraw from their temporary assignment and apply to $b(\sigma_b^T + 1)$. However, this contradicts the construction of the final DAwVW withdrawal step. □

The proof of DAwVW satisfying non-wastefulness follows along similar lines as property 4. Assume that there is such a "wasted" seat, such that there existed a school $b(\sigma^T)$ and student i such that i preferred $b(\sigma^T + 1)$ to the assignment they end up receiving. Moreover, suppose i is acceptable to b. For the algorithm to end, and for i to be assigned some outcome other than b, it would require i to choose not to withdraw and apply to $b(\sigma^T)$ in any step before the end of the final round. However, this contradicts the assumption that $b(\sigma^T + 1)$ is preferred by i, and the construction of the withdrawal step in the DAwVW algorithm.

One potentially additional interesting application of this property is that it might provide some testable implications of when a setting ought be modeled in the standard setup, compared to when the "size" of the resulting matchings should be included in the preference listing information elicited. Consider a standard school choice problem, where students are assigned according to either a TTC or SOSM (or some other stable mechanism). Given the results of this matching, ask if any student would like to join a school with an empty seat. If any student accepts this offer, it would provide direct evidence of the importance of school size.[64]

2.5.5 *Elimination of Extended Justified Envy*

Proposition 6. *The DA with Voluntary Withdrawals matching eliminates extended justified envy (EEJE).*

Proof. Showing the DAwVW algorithm's result satisfies extended EJE can be done most straightforwardly by considering the "at capacity school" and "below capacity school" cases individually.

First, assume the DAwVW result does not satisfy extended EJE because of a failure in case 1 from extended EJE's definition. This would mean there exists some student i who, at the conclusion of the DAwVW algorithm, is: matched to a school $b(x)$, acceptable to school c, c's final class size $y < q_c$ and $c(y+1) \succ b(x)$ for student i. However, this would contradict the necessary condition that the algorithm terminated, as i would have withdrawn from b before the conclusion of the algorithm. Moreover, if i withdrew from

[64]Note that a similar procedure could be considered for the results presented in Section 2.5.3.

b and applied to $c(y)$, given that $c(y+1)$ is more preferred by i than their final allocation $b(x)$, student i would not have voluntarily withdrawn from c in the subsequent withdrawal step. Thus, there can exist no such school c.

To prove that DAwVW satisfies the second case of the extended EJE definition, suppose DAwVW fails in this circumstance. Then, there must exist some student i who, at the conclusion of the DAwVW algorithm, is: matched to a school $b(x)$, acceptable to school c, c's final class size $y = q_c$, $c(y) \succ b(x)$ for student i, and there is some student j matched to $c(q_c)$ such that $i \blacktriangleright j$ for c. If i never applied to $c(q_c)$, it would contradict the necessary conditions for the algorithm to terminate, which itself contradicts the assumption that the algorithm terminated and yielded a matching. Alternatively, suppose i did apply to $c(q_c)$ in some step before the end of the procedure. If i applied to $c(q_c)$, it would mean $c(q_c) \succ d(z)$ for all other schools $d \neq c$. For i to end up matched to a school $b \neq c$, it must be the case that i was either rejected by c or "kicked out" of c due to a higher-priority student applying to $c(q_c)$. Since i is acceptable to c, it must be the case that i does not receive $c(q_c)$ because a higher priority student "took" the slot. However, by construction of the DAwVW algorithm, when there are more than q_c students applying for a seat in school c with maximum capacity q_c, the q_c highest priority students retain their seats at c. Since $i \blacktriangleright j$, it could not have been the case that j retained a seat while i did not. $\square$

2.5.6 Mechanism Is Not Strategy Proof

One unfortunate (though expected) result is that this mechanism doesn't satisfy strategy-proofness, which can be shown through a basic example:

Example 3 (DAwVW Not Strategy Proof)**.** Consider a problem with three students i, j, and k, two schools each with a capacity of two ($q_b = q_c = 2$).

Suppose the preferences are as follows:

$$i : b(1) \succ b(2) \succ c(1) \succ c(2)$$
$$j : b(1) \succ c(1) \succ b(2) \succ c(2)$$
$$k : b(1) \succ c(1) \succ b(2) \succ c(2)$$

And both schools have the same priority ordering of students:

$$b : i \blacktriangleright j \blacktriangleright k$$
$$c : i \blacktriangleright j \blacktriangleright k$$

The matching μ that corresponds to the outcome of the deferred acceptance with voluntary withdrawals in this case would be

$$\mu = \begin{pmatrix} \{i, k\} & j \\ b & c \end{pmatrix}$$

However, if student k manipulates their stated preferences, and instead reports:

$$\tilde{k} : c(1) \succ c(2) \succ b(1) \succ b(2)$$

The new outcome of the algorithm is:

$$\nu = \begin{pmatrix} \{i, j\} & k \\ b & c \end{pmatrix}$$

Where k has improved their allocation by misreporting preferences. ■

2.5.7 Limits on Coalition Manipulation

One additional result concerns the extent to which a coalition can benefit from manipulation. While a coalition of size n can, in principle, jointly manipulate their preferences to potentially mutually improve their outcome, the effectiveness of this manipulability is restricted in the same manner outlined for individual preference manipulation in Section 2.5.6. Specifically, these n students could "jointly play chicken" with those outside the coalition, however, like the individual case of manipulation in Section 2.5.6, the upper bound in terms of utility gained is one additional empty seat at their assigned school and comes with similar risks to the coalition (in terms of a possible worse outcome). To observe this limitation, consider the following two examples:

Example 4 (Successful Gain of One Seat)**.** Consider an augmented school choice problem using the DAwVW mechanism in round $t > n$. Suppose there exists some school x with n seats filled by a group of students acting as a coalition, and a maximum capacity $q_x > n + 2$. Moreover, suppose this n-student coalition's individual student preferences satisfy:

$$n_{\succ} : \; x(n) \succ y(r) \succ \ldots \succ x(n+1)$$

Where $y(r)$ represents some other feasible school-class size pair. Preference manipulation takes the form of all n students reporting:

$$\tilde{n_{\succ}} : \; x(n) \succ x(n+1) \succ y(r) \succ \ldots$$

Suppose some student i not in coalition n applies to $x(n)$ in a later round t'. If there exists some other non-struck option $y(r) \neq x(n+1)$ for student i, who has preferences:

$$i_{\succ} : \; x(n) \succ y(r) \succ \ldots \succ x(n+1)$$

Then student i voluntarily withdraws in t' instead of joining coalition n in school $x(n+1)$. ■

As long as the procedure terminates before any other student outside of coalition n applies to school $x(n+1)$, the coalition has successfully jointly "played chicken" with students outside of the coalition.

There are three forces, however, that mitigate practical ability for coalition members to successfully alter the matching outcome in this way. First is the fact that this requires a not-insignificant amount of information for the coalition members regarding preferences of other students. Second, if the coalition misjudges non-coalition preferences, it is possible that all members of the coalition end up worse off than if they had reported their preferences truthfully. Third, in cases where manipulation fails, there are individual-level incentives for members of the coalition to deviate from their joint strategy.

Example 5 (Unsuccessful Manipulation Attempt)**.** Consider an augmented school choice problem with the following preferences and priorities, and where students j and k are acting as a coalition, attempting to push student i to school b:

Note that, if student j were not part of the procedure, then the coalition's

True Student Preferences						
i :	$a(1)$	$a(2)$	$b(1)$	$a(3)$	$b(2)$	$a(4)$
j :	$b(1)$	$b(2)$	$b(3)$	$b(4)$	$a(1)$	$a(2)$
k :	$a(1)$	$a(2)$	$b(1)$	$b(2)$	$a(3)$	$a(4)$
m :	$a(1)$	$a(2)$	$b(1)$	$b(2)$	$a(3)$	$a(4)$
Coalition Manipulated Preferences						
$\tilde{k}$:	$a(1)$	$a(2)$	$a(3)$	$b(1)$	$b(2)$	$a(4)$
$\tilde{m}$:	$a(1)$	$a(2)$	$a(3)$	$b(1)$	$b(2)$	$a(4)$
School Priorities						
a :	i	j	k	m		
b :	j	i	k	m		

manipulation would be successful–declaring $a(3) \succ b(1)$ induces student i to withdraw from a and apply to b. However, adding student j (someone who prefers b at all class sizes to a) undermines this leverage. Now i is unable to get $b(1)$, and therefore applies to $a(3)$, where i joins k and m.

Critically, both k and m are worse off in this scenario than if they had truthfully reported $b(2) \succ a(3)$. This can be mitigated if either k or m had reported their preferences truthfully, however this would then undermine the coalition's leverage as a whole. ■

2.5.8 *Comparing Outcomes Between Standard DA and DA with Voluntary Withdrawals*

Before directly addressing Pareto efficiency, I describe several features of the outcomes derived from standard deferred acceptance algorithm and DA with voluntary withdrawals.

Claim 3. Fix a school choice problem. If all schools are at capacity for both the standard deferred acceptance and DAwVW algorithms, the two resulting matchings are identical.[65]

Proof. The proof follows directly from the process of "striking" schools from preference orderings in the adjusted problem's algorithm. Consider a matching where all schools are at capacity under the DAwVW algorithm. This necessarily implies that all matched students find at least one "at capacity" school acceptable, and that there are no "below-capacity" schools that are feasible (since, at the end of the algorithm, all schools are at capacity). By the "truncation step(s)" in each round, as outcomes become infeasible, those outcomes are removed from student preferences. Since only at-capacity schools remain at the end of round T, all under-capacity options have already been struck from student preferences. By assumption 3, the preference order of only at-capacity schools is analogous to the reported preference ordering in the equivalent standard problem. Finally, note that the two algorithms (standard deferred acceptance and deferred acceptance with voluntary withdrawals) both behave identically for at-capacity schools: students are only "dropped" from an at-capacity school if there exists a set of at least q students who prefer that outcome to all others. □

Claim 4. Suppose there exists a school b with maximum capacity q_b who, at the end of round t, is holding q_b students. No student in q_b will voluntarily withdraw from b in any future round.

[65] Intuitively, this highlights the fact that differences between the standard problem and the adjusted problem derive from the existence of empty seats.

Proof. The proof for claim 4 directly follows from the algorithm and from claim 2. Assume that claim 4 is false, and that a student was to withdraw from an at-capacity school in a round after it had reached capacity.

First, identify the most recent student in q_b who applied to b before b reached capacity, and denote this individual i. By assumption, every student holding a seat at b when i applies (as well as i themselves) must prefer $b(q_b)$ to any other feasible option. (Otherwise a student would withdraw, which contradicts the assumption that i was the final student to apply before b hit capacity.) Note that this also coincides with the conclusion of the round (since no student withdrew from b). By claim 2, in subsequent rounds, all other schools will weakly grow. Combined with assumption 2, this means all other feasible options for students holding an offer from b only get worse. However, the necessary condition for a student withdrawing from b is that there exists a more preferred feasible option, which contradicts the fact that no student previously withdrew when i applied.

The same steps can be taken to show that any student who applies to $b(q_b)$ after this point. Thus, contradicting the assumption claim 4 is false. □

Finally, the next proposition highlights a bound in terms of possible negative outcomes for individual students.

Proposition 7. *Any school assignment that is achievable under the standard Deferred Acceptance result is achievable under DAwVW.*

Note that proposition 7 does not suggest that the exact same outcome is achievable under both, however a student is guaranteed to receive at least the same school assignment.

Proof. Consider a school choice problem, and define μ^{DA} denotes the matching that results from the standard deferred acceptance algorithm and μ^{VW} denotes the matching from the DAwVW algorithm. Further, define the set of students matched school b in the DA and DAwVW matchings as σ_b^{DA} and σ_b^{VW} respectively.

Assume proposition 7 is false, and identify some student i matched to b in μ^{DA} but not in μ^{VW}. If $\mu^{VW}(i) \succ \mu^{DA}(i)$, the proposition is trivially true. Therefore assume $\mu^{DA}(i) \succ \mu^{VW}(i)$. If the school student i is matched to is not at capacity in either matching, then it is feasible (assuming the student is acceptable to the school).

The one condition not addressed by the two cases above is if the school i is matched to (say b) is at capacity under both μ^{DA} and μ^{VW}, and that i is matched to b under μ^{DA} but not under μ^{VW}. First, identify a student matched to b in μ^{DA} but not μ^{VW}, and denote this student j. Suppose j is matched to some alternative c under μ^{DA}. If c is not at capacity, then $c(q_c)$ is achievable for j under the DAwVW algorithm, which contradicts the assumption that the algorithm terminated (a necessary condition for the generation of μ^{VW}).[66] By construction of this case, since j is matched to b but i is not, and that i would prefer to be matched to b compared to their realized outcome under μ^{VW}, it must be the case that $j \blacktriangleright i$ at b.

In the DAwVW algorithm, it must have been the case that j lost their seat in c because of a blocking pair, (k, c). (In other words, student k preferred $c(q_c)$ to their assignment, and k was higher on the priority list of c than j.) However, this blocking pair would exist independent of the specific algorithm used: If k prefers $c(q_c)$ to all feasible options in the adjusted problem, then (by assumptions 3 and 2) they must prefer $c(q_c)$ to all other feasible options in the standard problem. Similarly, if k is higher on the priority list of c in the adjusted problem, they must be higher on the priority list in the standard one. This contradicts the fact that the standard deferred acceptance algorithm achieves stability by ensuring no blocking pair exists. $\square$

[66]Specifically, if j would rather c at a class size less than q_c, j would have withdrawn from b at some round in the DAwVW algorithm to apply to c. The fact that the algorithm ended implies this did not happen.

2.5.9 *Pareto Efficiency*

The first two results in this section reiterate circumstances previously identified. As a result, it is straightforward to observe that neither the standard deferred acceptance nor the DA with voluntary withdrawals uniformly Pareto dominates the other.

Claim 5. *The DA with Voluntary Withdrawals outcome does not Pareto dominate the standard DA matching.*

Proof. Claim 5 can be shown by a counterexample. Consider a case with four students $\{i,\ j,\ k,\ l\}$ and two schools b and c, both with maximum capacity $q_b = q_c = 3$. Furthermore, suppose student preferences are identical and satisfy the following:

$$\succ_{i,\ j,\ k,\ l}:\ b(1) \succ c(1) \succ b(2) \succ c(2) \succ b(3) \succ c(3)$$

And both schools share the priority order: $i \blacktriangleright j \blacktriangleright k \blacktriangleright l$. Finally, note that given these preferences, by Assumption 3, students would unanimously report $b \succ c$ in the equivalent standard problem.

Straightforwardly, the resulting deferred acceptance matching (when class size is not considered), denoted μ, and the matching that results from the DAwVW algorithm, denoted ν are:

$$\mu = \begin{pmatrix} \{i,\ j,\ k\} & l \\ b & c \end{pmatrix} \qquad\qquad \nu = \begin{pmatrix} \{i,\ j\} & \{k,\ l\} \\ b & c \end{pmatrix}$$

Although students $i,\ j,\ k$ all prefer ν to μ, student l prefers μ. □

Even more straightforwardly, the reverse can also be proven with an example employed earlier:

Claim 6. *The standard deferred acceptance algorithm, applied to the augmented school choice problem where students have preferences over class size, does not Pareto dominate the Deferred Acceptance with Voluntary Withdrawals result.*

Proof. Consider a scenario with two students i, j and two schools b, c with $q_b = q_c = 2$ (and both i, j are acceptable to both schools). Suppose student preferences are identical:

$$\succ_{i,\, j}: b(1) \succ c(1) \succ b(2) \succ c(2)$$

By Assumption 3, both students report $b \succ c$ in the standard deferred acceptance problem. The resulting deferred acceptance matching (when class size is not directly considered), denoted μ, and the matching that results from the DAwVW algorithm, denoted ν, are:

$$\mu = \begin{pmatrix} \{i,\, j\} & \emptyset \\ b & c \end{pmatrix} \qquad\qquad \nu = \begin{pmatrix} i & j \\ b & c \end{pmatrix}$$

Both i and j prefer ν to μ. $\square$

2.5.9.1 *Examples of Efficiency Gains with DA with Voluntary Withdrawals*

The lack of direct Pareto domination between the standard DA and DA with voluntary withdrawals outcomes does not mean DAwVW cannot lead to substantial welfare gains in various circumstances. In particular, as the following examples demonstrate two notions of efficiency gains in the extended framework:

- Welfare gains are not limited to low priority students. Even a student who is ranked first by all schools' priority lists, can see an improvement in the extended framework.
- Welfare gains are not limited to low priority students. Even a student who is ranked first by all schools' priority lists, can see an improvement in the extended framework.

To illustrate these situations, two examples are included below.

Example 6 (Inefficient Sorting Into Schools)**.** Consider a case with five students (i, j, k, m, p) and two schools (a, b), where both schools have maximum capacities of three students, $|q_a| = |q_b| = 3$. Suppose preferences and priority orders are defined accordingly:

i:	$a(1)$,	$b(1)$,	$a(2)$,	$b(2)$,	$a(3)$,	$b(3)$
j:	$a(1)$,	$b(1)$,	$a(2)$,	$b(2)$,	$a(3)$,	$b(3)$
k:	$a(1)$,	$a(2)$,	$a(3)$,	$b(1)$,	$b(2)$,	$b(3)$
m:	$a(1)$,	$a(2)$,	$a(3)$,	$b(1)$,	$b(2)$,	$b(3)$
p:	$a(1)$,	$a(2)$,	$a(3)$,	$b(1)$,	$b(2)$,	$b(3)$
a:	i,	j,	k,	m,	p,	
b:	i,	j,	k,	m,	p,	

Effectively, this describes a circumstance where all five students generally prefer school a to school b, however their relative sensitivity to class size differs significantly. Using the Consistency Assumption, the "standard problem" equivalent would be the case where all five students list $a \succ b$, which, combined with the unanimity in school-side priorities, would lead the SOSM, μ, to be:

$$\mu = \begin{pmatrix} \{i,\ j,\ k\} & \{m,\ p\} \\ a & b \end{pmatrix}$$

However, incorporating class size preferences, and allowing for voluntary withdrawals, the matching becomes:

$$\nu = \begin{pmatrix} \{k,\ m,\ p\} & \{i,\ j\} \\ a & b \end{pmatrix}$$

Which leads to an outcome where, for students i, j, m, p, $\nu \succ \mu$, while the final student, k is indifferent between the two. ■

A similar result can occur through a cascade effect—if one student, with a higher priority at a highly demanded school (one highly ranked on the preference list for more students than can attend), prefers lower ranked schools conditional on them being partially unfilled.

Example 7 (Cascade of Improvements)**.** Consider five students and three schools, where $|q_a| = |q_b| = |q_c| = 2$, and preferences can be described by:

i:	$a(1)$,	$b(1)$,	$c(1)$,	$a(2)$,	$b(2)$,	$c(2)$
j:	$a(1)$,	$a(2)$,	$b(1)$,	$b(2)$,	$c(1)$,	$c(2)$
k:	$a(1)$,	$a(2)$,	$b(1)$,	$b(2)$,	$c(1)$,	$c(2)$
m:	$a(1)$,	$a(2)$,	$b(1)$,	$b(2)$,	$c(1)$,	$c(2)$
p:	$a(1)$,	$a(2)$,	$b(1)$,	$b(2)$,	$c(1)$,	$c(2)$
a:	i,	j,	k,	m,	p,	
b:	i,	j,	k,	m,	p,	
c:	i,	j,	k,	m,	p,	

Now only the "top" student (who is unambiguously ranked ahead of all others by all three schools) is "class size sensitive," while the other four follow a more traditional lexicographic preference ranking. The standard SOSM result would be:

$$\mu = \begin{pmatrix} \{i,\ j\} & \{k,\ m\} & p \\ a & b & c \end{pmatrix}$$

However, allowing for voluntary withdrawals leads to an improvement:

$$\mu = \begin{pmatrix} \{j,\ k\} & \{m,\ p\} & i \\ a & b & c \end{pmatrix}$$

Where three students, k, p and i all improve, while the remaining two, j and m are both indifferent between μ and ν. ■

Examples (6) and (7) give cases where majorities of the five students are strictly better off under the alternative framework, while the others' welfare is neither improved nor harmed. However, it is possible to generate examples where all students strictly improve their allocation compared to the standard DA assignments (see appendix (A) for one such case).

2.6 Describing the Set of Matchings

2.6.1 Structure of Stable Set

Here, I briefly highlight a couple interesting features of the set of stable matchings for this type of problem. In particular, allowing for preferences over class size significantly relaxes the structure of this set. As described below, not only do these matchings no longer satisfy any lattice structure, but the number of students matched to schools is not constant across stable matchings.

For exposition, consider the following two school, five student example:

Student Preferences						
$i:$	$a(1)$	$a(2)$	$a(3)$	$b(1)$	$b(2)$	$b(3)$
$j:$	$b(1)$	$b(2)$	$b(3)$	$a(1)$	$a(2)$	$a(3)$
$k:$	$a(1)$	$b(1)$	$a(2)$	$b(2)$	$b(3)$	$a(3)$
$l:$	$a(1)$	$a(2)$	$b(1)$	$b(2)$	$b(3)$	$a(3)$
$m:$	$b(1)$	$a(1)$	$b(2)$	$a(2)$	$b(3)$	$a(3)$
School Priorities						
$a:$	i	j	k	m	l	
$b:$	i	k	m	l	j	

Suppose the maximum capacity of both schools is $q_a = q_b = 3$. There are four matchings that satisfy individual rationality, non-wastefulness, and elimination of extended justified envy, denoted μ, ν, η and ϵ below:

$$\mu : \left\{ \begin{pmatrix} i, k \\ a \end{pmatrix} \begin{pmatrix} j, l, m \\ b \end{pmatrix} \right\} \qquad \nu : \left\{ \begin{pmatrix} i, l \\ a \end{pmatrix} \begin{pmatrix} j, k, m \\ b \end{pmatrix} \right\}$$

$$\eta : \left\{ \begin{pmatrix} i, m \\ a \end{pmatrix} \begin{pmatrix} j, l, k \\ b \end{pmatrix} \right\} \qquad \epsilon : \left\{ \begin{pmatrix} i, j \\ a \end{pmatrix} \begin{pmatrix} k, l, m \\ b \end{pmatrix} \right\}$$

2.6.2 Non-Lattice Structure

One straightforward result of this example is that, unlike the "standard" problem without class-size considerations, stable matchings do not form a lattice in student preferences. While certain stable matchings can be generally more preferred to others (for example, students j and k both prefer μ to δ, while the other three students are indifferent between the two), there exists no maximally preferred element.

2.6.3 Assigned Students to Each School

Expanding preferences to class size also impacts classic finding that the set of students matched in every stable matching is the same (from Roth (1986)). Specifically, the number of students matched to a school under the augmented preference problem may be larger than the set matched under the traditional problem. The intuition is straightforward: there may be cases where a student finds partially, but not completely, filled schools acceptable.

Example 8 (Stable Matchings with Different Numbers of Students)**.** Consider the following example with three students and two schools, and capacities $q_a = q_b = 2$:

Student Preferences					
$i:$	$a(1)$	$b(1)$	$b(2)$	$a(2)$	$\emptyset$
$j:$	$a(1)$	$\emptyset$			
$k:$	$b(1)$	$b(2)$	$a(1)$	$a(2)$	$\emptyset$
School Priorities					
$a:$	i	j	k		
$b:$	i	j	k		

Which leads to two stable matchings, μ and ν:

$$\mu = \left\{ \begin{pmatrix} i \\ a \end{pmatrix} \begin{pmatrix} k \\ b \end{pmatrix} \begin{pmatrix} j \\ \emptyset \end{pmatrix} \right\} \quad \nu = \left\{ \begin{pmatrix} j \\ a \end{pmatrix} \begin{pmatrix} i, k \\ b \end{pmatrix} \right\}$$

It is straightforward to observe that both μ and ν satisfy the necessary components of stability: that no seat is wasted (i.e. there exists a student who would rather fill an empty seat than receive their allocation), there is no extended justified envy, and both are individually rational. At the same time, the set of matched students is quite obviously different—j is only matched to a school under ν, but voluntarily remains unmatched in μ. Not only is it possible for schools to have a different set of students matched under various stable matchings, the number of students matched itself may change.

One potentially informative inquiry is to examine what the matching would be under the "standard" problem without class sizes being incorporated: does ν represent an outcome with more students matched than the standard problem, or does μ reduce the total number of matches from the classical preference framework? "Recovering" the standard problem by applying the consistency assumption yields $b \succ a$ for students i and k, and $\emptyset \succ \ldots$ for student j. In other words, without allowing j to report preferences at the class-size level, they will simply prefer to sit out the procedure altogether. In fact, the SOSM for the standard problem is quite pronounced in its deviation from both μ and ν. Defining the standard SOSM $\mathcal{S}$:

$$\mathcal{S} = \left\{ \begin{pmatrix} \emptyset \\ a \end{pmatrix} \begin{pmatrix} i, k \\ b \end{pmatrix} \begin{pmatrix} j \\ \emptyset \end{pmatrix} \right\}$$

Where both μ and ν represent different Pareto improvements to $\mathcal{S}$.

2.7 Other Features of Note

2.7.1 Defining School and Class Size Capacities

One potential response to this idea from the perspective of a school system is to point out that, generally, schools are often at or close to capacity, which would substantially reduce any potential efficiency gains realized by adding the "voluntary withdrawal" feature to the standard deferred acceptance algorithm. While this extended procedure doesn't itself lead to welfare losses compared to a standard student proposing deferred acceptance algorithm, it is certainly possible that the added complexity might lead students to believe that preference manipulations are worthwhile, which could hurt the perceived legitimacy of the mechanism. However, even in cases where almost all schools are at their physical capacity, there are two scenarios that might still apply and lead to noticeable efficiency gains.

Case 1 (Gains From Individual Underdemanded Schools)**.** Even if there are only a handful of schools that are themselves underdemanded, if the number of open seats that are left unfilled at the termination of the utilized mechanism is itself unknown to all students before they are asked to rank schools, then students aren't even aware of the possibility of utility gains. For example, suppose a school is only three quarters filled in any stable matching outcome. Unless all students are not only aware of this when listing their preferences, but aware of the extent of this under-fill, and how all other students will react to this circumstance, students find themselves in a near intractable game problem. Due to the standard framework's incapability of considering the conditionality of students' preferences[67] students are forced to either "gamble" or "play it safe" in their stated preferences.

Case 2 (Efficiency Costs From Overcrowding)**.** Given that many systems operate in a scenario where several schools will necessarily be overcrowded, this procedure provides useful information about the costs of filling any school beyond its capacity. Since this procedure can elicit preferences over counterfactual school size ranges, it is possible to understand whether, given the reality that overcrowding can sometimes be a necessity, any alternative choice of student allocation would be a Pareto Improvement. This information may be critical for a publicly elected school board, since it best accounts for constituents' preferences.

Moreover, it is far more likely that at least one (if not both) of the above conditions apply for various school districts around the country. For example, consider the 2011-2012 Charlotte-Mecklenburg Schools and 2016-2017 Wake County Public School System reports regarding school utilization (In bibliography as Charlotte Mecklenburg Board of Education and CMS (2011) and WCPSS (2017), respectively).[68] In the CMS system, high school student utilization varies substantially, with eight of the 22 high schools in the

[67]That a student might only prefer this partially filled school to another, fully enrolled one *if* the former is sufficiently empty.

[68]These two were considered because of their status as large, but not overly, school systems, located in the same state (so utilization formulas are likely to be more standardized).

district at under 80% student utilization (that is, below 80% of the total capacity), and five of the 22 schools above 90% of the predetermined student capacity. Furthermore, of those five, two are technically overcrowded, with one at 113% student utilization.[69]

2.8 Conclusion

The primary goal of this paper is to highlight the possibility of improving outcomes in traditional matching problem frameworks by considering extensions that might more accurately represent the circumstances faced by market participants. Focusing on school choice problems, it is possible to see the substantial efficiency gains that are otherwise missed in the narrower "school-only" traditional framework. By extending the permissible preference structure reportable by students, "hidden utility" from emptier schools can be exploited. Assuming students have different sensitivities to school differences versus class size differences, allowing even an ordinal ranking of these deeper preferences can improve sizable numbers of students.

Although the mechanism presented in this paper is not strategy-proof (a feature that cannot be satisfied so long as rights are defined on only one dimension), the strict limit of how manipulable the mechanism is should alleviate some concerns about its implementation: students receive no worse of a school outcome than under the standard SOSM. Moreover, there are additional mitigating factors. First, it isn't entirely clear whether how manipulable the mechanism would be in large markets, given the relative rarity of the specific circumstances leading to successful preference manipulation. Second, while attempting to game the system by "playing chicken" with other potential students to get a smaller class might be successful in certain circumstances, doing so opens the manipulating student up to potentially substantial negative outcome effects: once the student commits to the false

[69]See page 139 of the report Charlotte Mecklenburg Board of Education and CMS (2011).

reporting strategy, they may well find themselves forced to pay the price and accept a full, (truthfully) less preferred option.

Chapter 3

Review of Experimental Economics Methodology

3.1 Introduction and Motivation

In this chapter, I seek to contextualize the results of "Compensation without Distortion: Stochastic Termination and Incentive Compatibility in Preference Elicitation" by reviewing methods, approaches, and findings from the broader experimental economics literature.

The rapid rise and proliferation of experimental methods across economics can be viewed in a number of different ways. Despite fewer experimental papers being published in top five economics journals in absolute terms, experimental work has, and continues to have, an outsized influence compared to other fields (Frechette et al. (2021), using citation counts). Experimental findings have directly led to a number of policy-relevant findings, including the design of FCC spectrum auctions (Cramton (1998)), default options (Dhingra et al. (2012)), environmental policy (Hahn and Metcalfe (2016) and Noussair and van Soest (2014) in addition to Friesen and Gangadharan (2013)'s review of experimental impact on pollution markets), incentive systems (Gneezy et al. (2011)), information on resource use (Allcott (2011)), tax evasion (Kleven et al. (2011)), mechanism and market design (Bolton and Ockenfels (2012) and Roth (2016)), among many others. Consider the impact of experimental analysis in labor economics. These tools have been used to test outcomes as far reaching as the impact of liquidity on labor market decisions (by examining the effects of bonuses for new Teach for America enrollees) in Coffman et al. (2019), educational externalities from school-based healthcare programs for children in Miguel and Kremer (2004), wage returns for years of schooling in Duflo (2001), minimum wage changes on employment and income (Card and Krueger (2000)), discrimination in negotiations, interview offers for jobs, and auditions (audit and correspondence studies by Ayres and Siegelman (1995), Bertrand and Mullainathan (2004), and Goldin and Rouse (2000) respectively) among many others.[70] It was the merger of theory with experiment that Robert Wilson identifies as the "missing ingredient" for market design to progress into a practical field (Roth and Wilson (2019)). Roth (1991) highlights the importance of field and laboratory studies in the development of empirical game theory. Perhaps it is little wonder that the experimental approach to economic analysis was behind the motivation for two of the three most recent (2019 and 2021) Nobel Memorial Prizes in Economics, and significantly influenced the work of numerous other Nobel Laureates.

In this chapter, my primary aim is to provide background into the field of experimental economics, paying particular attention to experiments examining individual preferences and relevant methodological critiques. This is not a definitive history of experimental methodology, nor a comprehensive review of literature in the field, of which there are several exceptional examples of both.[71] By primarily addressing decision experiments, I only make comparatively short reference to other experimental focuses and objectives, such as those examining strategic behavior or market formation and effectiveness. The restriction to a single type of experiment is driven by practical reasons. At the same time, it is worth mentioning that this restriction is not without precedent (see, for example, Levitt and List (2009)) nor theoretical reasoning. Game theoretic and market experiments not only seek to test different theories of behavior, they also utilize distinct mechanisms for eliciting information, asking questions, and compensating subjects. In section 3.3, I describe some of the history of experimental methodology (particularly as it relates to the forerunners of decision experiments). This is used to provide deeper context into the historical evolution of the field, and as a means to highlight recurring issues and concerns.

In addition to the theoretical "focus" of an experiment, other important

[70] For more on experiments in discrimination, see the comprehensive review Bertrand and Duflo (2016).

[71] Which I refer to whenever possible and appropriate.

characteristics are the setting, subject pool, and randomization mechanisms experiments utilize to gather data. Beyond *laboratory experiments,* which are conducted in highly controlled settings that largely attempt to abstract away from as many contextual signals (other than those being tested) as possible are a range of alternative collection protocols. Despite small differences in specific categorizations and definitions (for example, between methodological survey papers Harrison and List (2004) and Charness, Gneezy and Kuhn (2013)), important considerations are the nature of: methods of interaction, the subject pool and contextual information surrounding actions taken, interactions and motivations between subjects (both between each other and the experimenter), and the stakes of decisions. Following Charness, Gneezy and Kuhn (2013), *field experiments* are defined by underlying subject motivations: the actions being observed would have occurred absent the existence of an experiment.[72] A hybrid *extra-laboratory experiment,* defined by Charness and coauthors, can be thought of as a mixture of lab and field: having "the same spirit as laboratory experiments, but are conducted in a non-standard manner."

This last category of experiments can provide a significant amount of useful information, particularly when analyzing the external validity of laboratory data or the importance of context-specific information in behavior.[73] "Classic" applications of laboratory methodology in nonstandard environments can be particularly useful when attempting to generalize lab findings. Although far too numerous to cover here, examples range from using price lists to estimate differences in risk and time preferences between doctors and their patients in Athens, Greece (Galizzi et al. (2016)) to dictator and ultimatum games played by a nomadic tribe of hunter-gatherers in Tanzania (Marlowe (2004)) to measure altruism and cooperation.

Another use of extra-laboratory settings is in providing a novel setting

[72]This also captures *natural experiments*, which can be considered circumstances where a random process generates a controlled setting for analyzing models, however without any experimenter intent.

[73]See List (2007) and Gneezy and Imas (2017) for more discussion on this point.

for testing the effectiveness of laboratory methods and whether they capture real-world behavior. Results here have been somewhat mixed. For example, consider the question of how best to elicit risk preferences for rural farmers. Laboratory settings (and most standard economic theory) tend to favor incentivized mechanisms broadly akin to "multiple price lists" (first described in Holt and Laury (2002)). Some (for example, Jin et al. (2017)) find strong correlations between elicited risk preferences using multiple price lists and stated attitudes towards risk. Others, such as Bocquého et al. (2013) and Bougherara et al. (2017), use incentivized price list mechanisms to estimate risk and loss aversion parameters for farmers in France, finding the elicited preferences are consistent with previously observed "paradoxical" behaviors regarding the over-purchase of crop-loss insurance explored by Babcock (2015). However these results are not unanimously reinforced: Others (for example, Hellerstein et al. (2013) performed on Midwestern United States farmers) find such finely-tuned estimates of risk aversion are entirely uninformative for predicting an individual's insurance or crop-diversification decisions.[74] Finally, a notable body of literature has found difficulty applying these techniques to rural farmers in general: Galarza (2009), Cook et al. (2011), and Brunette and Ngouhouo-Poufoun (2019) all report varying, sizable proportions of their subjects had difficulty understanding the nature of the decisions posed (leading to apparent, significant intransitivities in lottery choice).[75]

The remainder of this chapter is as follows: sections 3.2 and 3.3 provide background and historical contexts regarding experimental methodology. Section 3.4 discusses several of the applications of decision-style experiments in the literature (including preferences towards risk in 3.4.1, ambiguity in 3.4.2, and altruism in 3.4.3). I then move to a brief survey of methods used to elicit preferences in section 3.5 (covering decision questions and price lists in 3.5.1, real effort tasks in 3.5.2, contests in 3.5.3, dictator games in

[74]This finding is echoes by Barham et al. (2014) who examine technological adoption rates and elicited risk preferences.

[75]It should be noted that, for this reason, Harrison et al. (2007) suggests experimenters add sufficient context to the questions asked in order to improve subject understanding.

3.5.4), and describe some of the implications of various procedures in section 3.5.5. Section 3.6 examines several of the payment mechanisms used in conjunction with elicitation methods, briefly describing surveys, pay all, and random selection mechanisms in sections 3.6.1, 3.6.2, and 3.6.3 accordingly, while I highlight several previously documented theoretical concerns with payment mechanisms in 3.6.4. Finally, alternative, less-frequently-used and more recently-developed procedures are detailed in section 3.7, such as paying a subset of participants in section 3.7.2, the PRINCE mechanism in 3.7.1, and the Accumulative Best Choice mechanism in 3.7.3.

3.2 Background

Numerous previous works have traced the history of experimental economics (notably including Roth (1993), Lee (2005) and European Conference on the History of Economics (5th : 1999 : Cachan, France) (2005), see also Chakravarty et al. (2011)). My intention in this section is not to echo those authoritative pieces. Instead, by drawing on these papers I hope to provide a consistent, relatively comprehensive description of how the earliest bricks in experimental papers led to the field as it exists today: Current methods, questions, and approaches are intertwined with these early works (and early criticisms).

Since their humble beginnings, experiments have proliferated across both economics specifically, and social sciences more broadly.[76] Laboratory settings have been effectively employed to better understand key elements of behavior, from informing new models of how individuals make decisions in strategic environments (see Camerer et al. (2004) and Crawford (1997)) and set expectations about future outcomes (Hommes (2013)). Observed patterns of altruism and generosity have influenced theory directly

[76]Reviews of, and guides for conducting, experiments can be found in law and economics Hoffman and Spitzer (1985), political economy Palfrey (2009), international relations McDermott (2011), moral and social philosophy Güth and Kliemt (2017), computer science Grossklags (2007), cognitive psychology Hertwig and Ortmann (2001), and others.

(see, among many others Rabin (1993)). It is because of its widening applicability and improvements in experimental methodology that in 2005, Larry Samuelson noted "experimental economics is currently making its transition from topic to tool" (Samuelson (2005), page 65), just as mathematical analysis did nearly a century ago.

The advantages of laboratory experimentation are straightforward, assuming necessary (and often difficult) aspects of their implementation can be satisfied. Laboratory experiments provide a standardized environment, well suited for testing theoretical predictions and encouraging replication.[77] Experiments provide "an important foundation for bridging economic theory and observation... allowing more direct tests of behavioral assumptions" (Davis and Holt (1992), 4). Moreover, these results are not narrowly confined to paradoxes or secondary observations. As Kessler and Vesterlund (2015) explain, even in circumstances where direct external validity cannot be clearly known, experiments provide useful arenas for identifying comparative static, directional, and qualitative results: providing insights akin to a wind tunnel's use in airplane design (Schram (2005)). Samuelson (2005) highlights the ability of well designed experimental tests can help fill important gaps regarding a theory's accuracy, informativeness, precision, and usefulness. In some circumstances, this can take the form of "debunking" theories based on observed behavior. However experimental tests are not limited to 'shooting-down' the results of other fields, Samuelson and Binmore (1999) also emphasize the role of experiments in assisting with improvements to theoretical models that can be 'exported' to other arenas.

There are several prominent theories of decision making that rely heavily on empirical patterns, often discovered (or bolstered) through experimentation. Even though Prospect Theory's roots in Kahneman and Tversky (1979) might not meet the traditional 'economic experiment gold-standard' standard currently in use, Cumulative Prospect Theory (CPT) has both proved useful in decision modeling and had many of its central observations repli-

[77]See, for example, Croson (2002)'s discussion.

cated in more rigorous environments.[78] Even behavioral decision theories developed independent of experimental evidence have been tested and refined through testing. Yaari (1987)'s "Dual Theory" makes only indirect reference to basic experimental results, which is more than Sims (2003) in developing a theory of Rational Inattention. However both have been refined, at least partially due to their exposure to laboratory testing. In GURIEV (2001)'s generalization of Dual Theory, Guriev points out the surprising inaccuracies produced by Yaari (1987)'s model and prior empirical results (pointing to works like Hey and Orme (1994)). Further, Cox et al. (2012) constructs an experimental environment that allows for calibration tests (a la Rabin (2000), Safra and Segal (2008), and Sadiraj (2013)) of dual theories of utility. More prominently, rational inattention has been observed, replicated, and extended based on laboratory findings. For example, Caplin and Dean (2015) develop a test intended to differentiate "mistakes" from costly information acquisition using data that can "be readily gathered" in the lab. Others, such as Geng (2016) and Hebert and Woodford (2019), use experimental data about decision times to inform and refine models of rational inattention and costly information acquisition.

In other situations, the creation of a theoretical model and the development of an environment in which to experimentally test it fit uniquely well. In this space rests works such as Gneezy et al. (2003), who simultaneously develop a model of differential preferences by gender over competition and a novel experimental protocol in order to test those preferences.[79] Alternatively, Miller (1984) proposes an "Item Count Technique" method in order to simultaneously measure stigmatization and review the effectiveness of

[78] Namely, Kahneman and Tversky (1979) asked unincentivized survey questions to subjects, while the modern "gold standard" is to incentivize subject responses directly through compensation (see, for example Tversky and Kahneman (1992) and, more recently, Barberis (2013) for surveys; Gächter et al. (2021)). It should be noted, however, that tests of CPT do not unanimously support its modeling approach compared to other non-linear probability-weighting utility theories like Rank Dependent Utility Theory introduced by Quiggin (1982) (on this point, see Harrison and Swarthout (2016)). Still others dispute the evidence for behaviors predicted by CPT such as loss aversion (for example, Ert and Erev (2011) and Walasek and Stewart (2015)).

[79] The mechanism developed being so useful, it has been employed to measure behavioral preferences towards competition in other settings as well.

standard survey instruments. While initially applied to measure the prevalence of stigmatized behaviors directly (see Imai (2011) for a review and discussion), Coffman et al. (2017) demonstrate its effectiveness in a wider variety of settings, including measuring sentiments, attitudes, and biases held against stigmatized populations. The need for simple and straightforward measures of attitudes towards uncertain prospects led to Dimmock et al. (2016)'s generalized measure of ambiguity attitudes—a result subsequently used in its own right in subsequent theoretical models.

One particularly significant concern is of subject motivation: one of the cornerstones of experiments in economics–as opposed to psychology and other social sciences–is the focus on the importance of incentives created by an experimental design. Davis and Holt (1992) describe the two aspects of salient motivation: "(1) that the subjects perceive the relationship between decisions made and payoff outcomes, and (2) that the induced rewards are high enough to matter in the sense that they dominate subjective costs of making decisions" (p. 24).[80] Despite some minor disagreement on the extent of compensation effects, meta-analyses and reviews have largely supported the importance of salient and sufficient compensation in experimental results, particularly as it relates to decision experiments (with the most widely referenced being Camerer and Hogarth (1999)).[81]

Closely related to the question of *how much subjects should be paid* is *how those payments should be determined?* The specific concerns relating to this issue is domain and context dependent, however has nonetheless been raised in several prominent theoretical works over the past decades. A collection of four papers–Cox et al. (2015), Harrison and Swarthout (2014), Charness et al. (2016), Azrieli et al. (2018)–all provide detailed theoretical analyses of experimental payment mechanisms used in contemporary decision the-

[80] Despite some notable analyses of the cognitive-cost aspect of compensation (for example Smith and Walker (1993)), overall, the former has received more attention than the latter.

[81] A closely related issue in experimental methodology is the effect of costly effort on behavior. See Charness et al. (2018) for a review on techniques and theoretical considerations.

ories.[82]

3.3 History of Experimental Economics Payment Methods

Some of the first mentions of "experiments" in mainstream economics refer to a hypothetical or "metaphorical" investigation of a representative agent's behavior.[83] While interesting, these early works are more akin to (now ubiquitous) appeals to economic intuition than an attempt to gather and analyze individual-level behavior.

Initial "real" attempts to perform experimental analysis tended to use hypothetical choices. Thurstone (1931) asks subjects to choose between sequences of bundles consisting of overcoats, hats, and shoes as a way to measure stated preferences and derive indifference curves. Closely related in time and incentive design, Gilboy (1932) collects and uses survey data (asking about both real-world and hypothetical-choice scenarios) to "relate [previously observed] empirical curves to" more formal functional theories of decision making.

Critiques of these early experimental steps tended to focus on methodological concerns (though often in a general sense). Representative of the response is Georgescu-Roegen (1936) who, despite conceivably useful results, fears problems in application and a lack of sufficient formal rigor are likely to doom the usefulness of experiments.[84] Wallis and Friedman (1942)'s critique highlighted a number of methodological issues with Thurstone (1931), such as the importance of a stationary decision-environment and the need for a generalized setting free from restrictive impositions on behavior. Two more fundamental problems undermined the perceived usefulness of experimentation in economic settings more generally. First was the imposition

[82] The first two also provide evidence regarding biases induced by payment mechanisms in decision settings, which is further described in sections 3.5 and 3.6.4.

[83] See Lenfant (2012)'s references to Fisher (1892) and Vilfredo Pareto's writings.

[84] As Lenfant (2012) quotes Georgescu-Roegen (1936): "The method of economics remains—and it seems that it will remain despite many attempts in the opposite direction—that of the mental experiment aided by introspection."

of hyperbolic indifference curves; as a result, it becomes impossible to identify and account for subject behavior that systematically deviates from this strict functional form. Second was the hypothetical nature of the questions posed. Wallis and Friedman (1942) argue that theories describe behavior in real scenarios, facing actual consequences, and "hypothetical stimuli do not satisfy this requirement. The responses are valueless because the subject cannot know how he would react."[85]

The "Wallis-Friedman" critique of Thurstone's hypothetical survey design was, as Roth (1993) describes, challenged head on by Rousseas and Hart (1951), which lays out a more direct approach: "In principle, it would be possible to pu a test individual through a sequence of concrete choice situations, which would tell us a good many characteristics of his preference map—provided that the test was short enough in time to lend credence to the assumption that his preferences were constant over the duration of the experiment." Although the setting ultimately utilized by Rousseas and Hart (1951) is subject to its own pitfalls (Columbia university graduate students being asked to rank three potential breakfast "bundles" comprised of mixtures of bacon and scrambled eggs), it nonetheless represents a significant step in the evolution of experimental design.[86]

Most of the earliest incentivized experiments are closest mechanically to versions of "pay all' designs, where every decision a subject made was directly compensated. Mosteller and Nogee (1951) performed one of the earliest "modern" discrete choice experiments by asking subjects to make a series of decisions over monetary gambles and basing compensation on elicited preferences. Although there are, of course, additional steps between this innovation and modern methodological critiques, I leave that discussion for other authors.[87]

[85] Even authors who were more sympathetic to the underlying goal of understanding the existence and nature of indifference curves nevertheless rejected the role of experimentation in their derivation. See Lenfant (2012)'s discussion of Samuelson (1950).

[86] Although, as Roth (1993) argues, it is possible that this was framed as an incentivized experiment, but the students were not actually fed their most preferred choice.

[87] For more discussion, see the general history of experimental economics papers referenced at the start of this section.

A second contemporaneous strand of literature would similarly be recognizable today: giving subjects a series of choices and estimating which utility theories are best able to "fit" elicited responses. Coombs et al. (1967) conducted two experiments–one hypothetical, one real–asking subjects to make choices between 47 pairs of gambles.[88] This goal and general approach is echoed in modern experiments, particularly those interested in estimating probability weighting functional forms and comparing expected against non-expected utility theories.[89] While these are instrumental for informing proper model specification in non-experimental contexts, others have urged caution in their interpretation. Critiques range from misspecification of elicitation measures (such as Walasek and Stewart (2021)'s critique for testing prospect theory) to broader concerns over the implicit incorporation of functional forms in parameter estimation (see, for example, Torres et al. (2011) and Stewart et al. (2019)).

3.4 Uses of Decision Experiments

Eliciting preferences regarding bundles of goods, states of the world, or beliefs about future outcomes are instrumental in providing a useful view into revealed preferences. By identifying points of disagreement between theories, or gathering specific information about behavior, strengths, weaknesses, and limitations of competing models of behavior and choice can be compared. Decision experiments provide a unique method of observing and measuring those particular elements of behavior that are otherwise unavailable.

As described more generally above, one concern is that experimentally elicited preferences may not generalize to other settings. I thus try to discuss evidence of external validity in each of the following settings.

[88]One distinction between this procedure and modern methods: the real experiment included gambles over cigarette allotments instead of money directly.

[89]For examples, see Gonzalez and Wu (1999) and Wilcox (2015) among others. For a more complete discussion of the modern approaches to estimating probability weighting, see Booij et al. (2009).

3.4.1 Risk

One of the largest bodies of work in decision experiments examines preferences over known-probability, risky outcomes. As these results comprise the plurality of lab experiments described in this chapter (and the majority in the "Compensation without Distortion" chapter), I will not spend too much time rehashing the results. In addition to the tools described in section 3.5, experimentalists have utilized portfolio-choice tasks, which ask participants to allocate an endowed "budget" over a constructed menu of risky "assets."[90] The mechanism chosen is not innocuous; eliciting preferences over portfolios, compared to binary choice or price list methods, fundamentally change the nature and framing of the experiment. On one hand, this can make risk estimation more difficult, as it is challenging econometrically and theoretically to disentangle multiple simultaneous effects. On the other, the underlying structure of the task is more intuitive, and more closely matches real-world experience, than picking between two contextless lotteries.

Attempts to generalize risk preferences from laboratory data to behavior has been decidedly mixed. Dohmen, Huffman, Schupp, Falk, Sunde and Wagner (2011) performs an extensive analysis, and does find elicited risk preferences from a lottery choice task do correlate with an individual's willingness to invest, and with health and occupational decisions. More recent results from Charness et al. (2020) arrive at the opposite conclusion: that while risk elicitation methods are correlated with risk environments induced in a lab (for example, in portfolio or investment experiments), they do not predict real world behavior such as investing in stocks or likelihood of having a savings account.

One likely shortcoming of these analyses is the inability to disentangle risk as elicited from various mechanisms from functional assumptions over

[90]For a comprehensive analysis of these procedures, and how they compare, see Holt and Laury (2014). Several notable portfolio-choice experiments on risk and preference consistency include Gneezy and Potters (1997), Loomes and Segal (1994), and Choi et al. (2007).

utility. For example, Charness et al. (2020) use elicited responses to estimate the risk parameter of a standard CRRA utility function. Of course, if this function is itself misspecified, it isn't entirely clear what the expected result of the analysis is (a point made in their conclusion as well). Another issue is the framing of experimental payments in subjects' minds—see Hvide et al. (2019) as a recent example of the importance of "windfall" vs. earned income framing in experiments.

For additional resources regarding measurement of risk preferences in a laboratory setting, see Elliott (1998), Charness, Gneezy and Imas (2013), and for a critical reanalysis of the impact of incentives on elicited risk preferences, see Eckel (2019). For a review of eliciting risk preferences outside of a lab setting, see Barseghyan et al. (2018) and Charness and Viceisza (2016). It is also worth noting the importance of theoretical and experimental congruence when estimating risk preferences; O'Donoghue and Somerville (2018) provides a useful discussion of how to model behavior surrounding risky choices.

3.4.2 Ambiguity

One body of experiments examines the effect of ambiguity on decision making. Unlike circumstances where the probabilities of outcomes are known to the subject, ambiguity (also called *Knightean Uncertainty*, following Knight (1921)) seeks to understand behavior in cases where these probabilities are unknown. One of the most famous paradoxes demonstrating a failure of Expected Utility Theory, the Ellsberg Paradox from Ellsberg (1961), rests on the importance of ambiguity for decision makers. This importance has implications beyond tests for Expected Utility, preferences over ambiguous bets have been addressed directly by decision models (see, for example, Dillenberger and Segal (2017)), other "non-standard decision behavior" (Dean and Ortoleva (2019)), neurobehavioral studies (for example, Wu et al. (2021)), among others. For a more comprehensive examination of these additional implications of ambiguity preferences, see Bühren et al. (2021).

Chew et al. (2017) construct a series of decisions under "partial ambiguity," where subjects make choices where the probability of winning is determined by a disjointed, but symmetric, distribution of probabilities. Cubitt et al. (2018) calculate ambiguity and risk premia for subjects by eliciting preferences between a known outcome and either a risky bet, or an ambiguous one (where the ambiguous-bet probability is determined by a separate stochastic process), and finding ambiguity and risk premia for are of similar magnitudes. Baillon et al. (2018) estimate subject ambiguity aversion and sensitivity parameters by asking about preferences between an uncertain event (the change in the Amsterdam Stock Exchange between the start of the experiment and the end of a predefined time interval) and a risky one (a gamble between known prospects) and the effect of time pressure on these parameters. Others, such as Ahn et al. (2014) have turned towards portfolio choice settings (similar to the procedure mentioned in the previous section), who generally construct environments that allow subjects to construct optimal portfolios of invented assets, where different options yield differential exposure to ambiguity in either outcome or probabilities of outcomes.

One aspect of note is the frequency of random incentive systems in experiments on ambiguity. The effect of assuming, often implicitly, some version of the independence axiom on preferences for ambiguity is not necessarily clear.[91] Even in cases where this randomization device is used, there is often little or no analysis of the implications of this assumption (see Yang and Yao (2017) as an example). Others, such as Echenique et al. (2019), describe the use of a version of the random incentive system, though without a formal analysis regarding the potential distortionary effects of this payment scheme.[92]

[91] It is arguable that findings in some of these experiments, at least weakly, suggest assuming the necessary form of independence violates the preferences for at least some decision makers. For example, Baillon et al. (2018) note that time pressure increases violations of set-monotonicity in their results.

[92] Technically, Echenique et al. (2019) use a method closer to the PRINCE mechanism, as described in section 3.7.1.

3.4.3 *Altruism and Social Preferences*

In other settings, dictator games have been used as instruments for measuring baseline altruistic preferences to explain other observed behavior. For example, Dreber et al. (2014) use donations in a dictator game to demonstrate cooperation in repeated prisoner's dilemma games is motivated by selfish, and not altruistic, preferences.

At the same time, others have used variations of dictator games to argue against the relative importance of social preferences. More specifically, relaxing the decision environment (along the lines as those discussed in section 3.5.4) can alter the behavior of givers, suggesting significant experimenter demand effects. Other environments used to measure altruistic behavior in the lab, such as public goods games, trust games, and repeated prisoner's dilemma games, have yielded evidence of altruism (see the discussion in Andreoni et al. (2010)).

Alternative approaches attempt to mitigate the "sterile" and contextless decision environment created in a lab setting to, perhaps, more accurately measure altruism as it exists in society. Some examples of this method include Andreoni et al. (2017) (who test the effect of explicitly asking for donations among Salvation Army bell-ringers), List and Lucking-Reiley (2002) (examining the effect of seed money on donation solicitation mailers), and Mujcic and Frijters (2011) (observing behavior of commuters at an intersection). While these are likely to avoid a good amount of problematic demand effects present in a laboratory environment, these results are also more difficult to generalize and for testing theoretical implications.

3.5 Elicitation Procedures

Knowing *what the experiment is seeking to test* is the first step. Of course, once this is resolved, the next question that must be answered is *what data* *are needed to test the hypothesis?* Data requirements and context are inexorably linked to an experiment's design most directly through the choice of elicitation procedure. An experiment's elicitation procedure can effectively be considered its user interface; subjects are (likely) unaware of the hypothesis being tested, and might not even fully understand how their behavior impacts their compensation (if it does at all).However every lab experiment must have some way for individuals to be able to interact with the experimenter.

Informally, the elicitation method answers the following questions:

1. What types of questions will be asked to participants?

2. How are those questions asked?

3. How do subjects answer the questions that are asked to them?

This section provides a bit of detail regarding several methods used in decision experiment environments.[93]

3.5.1 Choice Questions

The backbone of experimental economic methodology is no doubt the choice question. At it's heart, a choice question simply asks a subject which option (or what valuation, or sequence of options, or beliefs, etc.) best fits their preference structure. In experiments attempting to elicit risk or ambiguity preferences, choice questions often take the form of 'which of the following options do you most prefer?' It is, effectively, the experimental equivalent of a direct elicitation mechanism.

[93]This is not intended to be an exhaustive review, but rather an informative discussion. For more of the former, see Charness, Gneezy and Imas (2013)'s analysis of elicitation procedures for measuring risk. Experimental investigations of the effects of elicitation procedure on observed behavior in different contexts include Freeman et al. (2016), Holzmeister and Stefan (2020) and Bénabou et al. (2020).

The advantages of this approach are straightforward and, if the questions are clear, hypothesis is well defined, and incentives are aligned, produces the most straightforward data for testing theories. Difficulty can quickly arise, however, in contexts where the links between these three elements becomes tenuous. For this reason, and a need to standardize best practices, a growing literature has sought to develop recommendations and publicize shortcomings of choice questions when applied to particular fields. While this effort is perhaps most well defined in health economic experiments (see, for example, de Bekker-Grob et al. (2010), Vass et al. (2017) Hauber et al. (2016) for discussions of a variety of methodological issues in health decision modeling), though is ongoing in other fields, such as resource economics (Johnston et al. (2017)) and the social sciences more broadly (Atkinson (2015)).

3.5.1.1 Multiple Price Lists

A distinct subset of choice questions follows from the instrument developed by Holt and Laury (2002), interchangeably referred to as the **multiple price list** (MPL) and **Holt-Laury** mechanism. At its heart, the MPL presents subjects a sequence of binary choices at once. Traditionally, one column presents a uniformly riskier option than the other. At one end (either the top or bottom of the list), the riskier column yields a significantly higher payoff, while at the opposite end, the safe option has a higher expected value. Every row in the multiple price list then offers an incremental step along this chain. By observing where an individual switches from preferring the safe to risky options, the experimenter is able to estimate a bracketed parameter estimating the subject's risk preferences.[94]

Multiple Price Lists have been used in a range of settings for measuring general risk aversion, such as in health and risk behavior (Anderson and

[94] It is also customary to pay subjects first by choosing a random row and basing payment off the choice the subject made in that specific circumstance, as Holt and Laury (2002) do. Effectively, the entire price list becomes a random incentive system, described in more detail in section 3.6.3.

Mellor (2008)), conjoint risk and time preferences (Andersen et al. (2008), also see Cohen et al. (2020) for a more thorough discussion of time preference estimation), choice architecture (Allcott and Kessler (2019)), and in conjunction with measurements of job choice and cognition (Burks et al. (2009)). Of course, these applications are not without potential drawbacks. One is the direct equivalence between a multiple price list and a random incentive system, and the associated incentive implications for choice distortions (see section 3.6.4). In addition to these mechanical concerns, Brown and Healy (2018) demonstrate the existence of a "list framing effect," where subjects are behaviorally discouraged from choosing options at either end of the list (and therefore tend to "hedge" by choosing a middle-option). This builds off a larger (related) body of framing effects more generally, which can also impact choice.[95]

3.5.2 Effort Tasks

Several modern theories of decisions incorporate the underlying idea that making choices is often difficult. In response, experimentalists have developed several tools to induce effort into experimental settings.[96] Gächter et al. (2015) devise an electronic "ball catching" task, where a subject moves a "bucket" at the bottom of the screen in order to be underneath simulated balls as they fall from the top of the screen.

In certain contexts, incorporating mental or physical effort is critical for hypothesis testing (either for confirmation or rejection of an underlying theory). Surveys of experiments that induce subjects to put in effort (see Deck and Jahedi (2015), for example) have found replicated evidence of cognitive effort's effect on perceptions of characteristics such as risk.[97] At the same time, other settings have demonstrated active subject preferences for tasks

[95]For a more thorough discussion of this broader concern, see Andersen et al. (2006).

[96]For a more thorough review of these tasks, see Carpenter and Huet-Vaughn (2019).

[97]Somewhat similarly, Filippin and Gioia (2018) find that men's risk preferences are significantly impacted by their performance in a competitive environment, while women's risk tolerance is unaffected.

that either add variety, ensure future flexibility, or reduce cognitive load (see, for example, Dean and McNeil (2014), Schouppe et al. (2014)).

3.5.3 Contests

If the researcher is interested in understanding behavioral characteristics, such as the effects of competition (or its avoidance), it is often necessary to reflect at least some aspect of competition in the tasks directly. Gneezy et al. (2003), for example, demonstrate that the existence of competition can impact performance (with the finding echoed by Gneezy and Rustichini (2004), who use an "extra-lab" design by observing the impact of competition on Israeli elementary-school aged children's running speed).

Alternatively, one could primarily be interested in the effects (or existence) of *preferences* toward competition. Some of the earliest experiments on this front tested differential competitive preferences across gender. Niederle and Vesterlund (2007) propose one of the foundational elicitation methods for competitive preferences: competition is measured by behavior in four stages, all of which require the subject to complete as many addition problems (each consisting of four numbers, each with two digits) as possible in a five minute period.[98] The first two stages acclimate the subject to the two possible conditions–in the first, they are paid a rate of $0.50 per correct answer, while in the second, they are paid based jointly on correct answers and whether they answered more than three other participants (randomly selected to create groups of four). In the third and fourth stages, participants may choose either to compete against their group's (lagged) scores or return to the constant-rate payment system.[99]

[98] Villeval et al. (2005) deserve a mention here, partially bridging the methods of the two other major papers: using the maze-solving tasks from Gneezy et al. (2003), but allowing individuals to choose their compensation scheme, somewhat similar to Niederle and Vesterlund (2007).

[99] The rules regarding what one's score is compared to differs between the third and fourth round, allowing Niederle and Vesterlund (2007) to decompose preferences for competition from the effects of competition on performance.

Since the demonstration of Niederle and Vesterlund (2007)'s mechanism's effectiveness, others expanded this approach in several directions. One set of extensions attempts to more directly identify the root cause of gender differences in preferences. Gneezy et al. (2009) compare competitive preferences between a strongly patriarchal (the Maasai in Tanzania) and matriarchal (the Khasi in India) society, finding the dominant gender in both cases demonstrates stronger preferences towards competition. In other cases, the competition-elicitation tool was used to examine differences in preferences between alternative populations. For example, Charness and Villeval (2009) demonstrate generational differences in willingness to compete between "junior" employees (younger than 30) and "senior" employees (older than 50) of manufacturing companies in Lyons, France.

As the use of competition elicitation procedures has expanded, the results are not necessarily monotonic and entirely consistent, suggesting the importance of related characteristics like culture and social interactions. For example, Zhang (2013) measures preferences for competition in rural middle school-aged Chinese students.[100] Zhang finds no gender effects, however observes more competitively inclined students are more likely to choose to take a potentially high-reward (though costly) state-issued exam after controlling for scholastic ability.

Some who have analyzed group differences across types of competition tasks do recommend care is taken to model the specific type of competition correctly, and warn against relying too heavily on a single "competition preference" parameter. Lezzi et al. (2015) perform a series of experiments where treatment groups differ by the nature of the competition faced by participants.[101] Their results yield little measured consistency between different treatment arms.

[100]The design is similar to Niederle and Vesterlund (2007): students are asked to add four two-digit numbers together in four stages. The stage of particular interest is the third, where students can either be paid based on their absolute performance, or their relative performance compared to a group of their classmates.

[101]More specifically, competition between subjects was conducted by asking individuals to adjust sliders to a particular position on a screen, answer math questions (sum of two, two-digit numbers), count "1"s in a grid of "1"s and "0"s, and an investment task.

3.5.4 Dictator Games

One of the most commonly used environments in decision experiments is the "dictator game," first as a three-player, hypothetical game introduced in Kahneman et al. (1986) and later modified to the more ubiquitous two-player real stakes experiment in Forsythe et al. (1994). Effectively, the dictator game represents a "simplification" of the ultimatum game, removing any element of interaction (transforming the setting from a game-theoretic to decision environment). While there are nominally two "players," payoffs are determined solely through one player's actions. In short, one individual is granted an endowment by the experimenter, and is asked how much they would like to donate to the non-endowed subject.[102]

In the 28 years since Forsythe et al. (1994), there has been a significant body of literature using the dictator game environment to test a wide variety of hypotheses.[103] It has been used, in one form or another, to elicit preferences for fairness and reciprocity (see, for example, Charness and Rabin (2002)), justice (Schurter and Wilson (2009)), and others' welfare (García-Gallego et al. (2019), among many others). Moreover, by altering the structure, previous authors have expanded the mechanism's use. Instructive here are Felix and Reiner (2008), who ask senders to allocate tickets to a raffle with a monetary prize (instead of a sum of money directly) to measure preferences in a probabilistic setting, and Dana et al. (2006*b*), who allow dictators to "pay" the experimenter in order to avoid telling the receiver that the dictator game was even being played at all.

One major concern regarding dictator game environments for estimating general preferences for altruism is their apparent sensitivity to experimental design choices. The relative level of experimental "blinding" has

[102]The idea of removing the ultimatum game receiver's ability to reject offers had been used in related, though more complex settings before Forsythe et al. (1994). See Camerer and Thaler (1995) for an early review of these early variants.

[103]A search for 'Dictator Game' in RePEc abstracts between 1995 and 2022 yields over 1,000 results.

been shown, though to varying degrees, to impact giving behavior.[104] More directly, several notable findings yield evidence that dictator donations are themselves driven by social expectations or experimenter demand effects, instead of personal feelings of altruism. Bardsley (2007) and List (2007) relax the environment, allowing dictators to either give or take from a recipient, and both subsequently report a significant decrease in the amount and frequency of positive transfers to recipients.[105]

3.5.5 *Importance of Elicitation Procedure*

Numerous early works identify the theoretical importance behind the choice of elicitation method. Karni and Safra (1987) demonstrate that a BDM mechanism can itself induce a subject's preferences to become "distorted" for non-expected utility maximizers.[106] Beyond these theoretical concerns, experimental evidence similarly suggests that the non-intuitive and cognitively complex mechanism can impact stated preferences (see, for example Predmore et al. (2021)). Several papers have more generally sought to estimate the effects of elicitation procedure on revealed preferences. Findings of this group have yielded significant, persistent procedure-specific distortions in contexts ranging from time preferences (Freeman et al. (2016)), to risk (Holzmeister and Stefan (2020)) to moral and ethical contexts (Bénabou et al. (2020)). Although all the above demonstrate the method-specific effects of various procedures, solutions to these observed violations of procedure equivalence are far more difficult to come by.

There are other circumstances, however, where some strongly suggest

[104] Hoffman et al. (1996)'s early finding, that giving decreases as social distance from the recipient increases has been replicated by others, for example Koch and Normann (2008), though of a lesser magnitude. It should also be mentioned that part of the decrease in giving could be driven by senders' lack of belief in the fidelity of the experiment; as Frohlich et al. (2001) describe, as social distance between the sender and receiver increases, senders may begin to believe the opposite party does not, in fact, exist.

[105] At the same time, both also report a relative unwillingness among dictators to simply choose the action with the highest personal payoff, suggesting a notion of fairness inconsistent with a standard rational choice theory approach.

[106] A result expanded upon in Safra et al. (1990) and generalized in Horowitz (2006).

emphasis on differences between elicitation mechanisms is misplaced. Returning to preferences over competition, several recent analyses find an equivalence between revealed preferences in the vein of Niederle and Vesterlund (2007) and standard "psychometric scale" question responses (where individuals are simply asked to rate, on a given scale, how competitive they are).[107] Do these results mean that there is no reason to use incentivized contests? Not necessarily, but (assuming the equivalence results do hold generally) there may be a strong case to be made in favor of experiment-dependent elicitation choice. If the existence of a well-calibrated multi-round competition is likely to impact performance in another task, there may not be too much lost by utilizing a survey question to elicit competitive preferences.

3.6 Payment Procedures

Metaphorically, the payment mechanism is the engine that drives the experiment: it describes the conditions in which choice questions are asked, impacts how the elicitation procedure is structured, and determines how subjects' payments are generated. The ultimate objective behind tying subject compensation to actions taken is to ensure *incentive compatibility*–effectively a way to align the participants' incentives with those of the experimenter. For a thorough theoretical treatment of payment procedures and their impacts on incentives, see Azrieli et al. (2018). Alternatively, Cox et al. (2015) and Harrison and Swarthout (2014) analyze how the choice of payment procedure impacts experiments measuring risk preferences, each testing some of the theoretical concerns raised through their own experimental analyses. In this section, I introduce several generalized forms of the most common methods used to compensate participants, and briefly touch on known advantages and disadvantages. More recent papers have developed new ways

[107] See, for example, Bönte et al. (2017) and Fallucchi et al. (2021), who find a strong relationship between both measures of competitiveness.

of measuring preferences and compensating subjects—for a review of several of these alternatives, see section 3.7.

3.6.1 Surveys

The unifying feature of survey instruments is their disconnection between answers subjects provide and their compensation. (As described in section 3.3, this approach covers the majority of early experiments in the field.) More recently, surveys have been used both in place of incentive mechanisms (effectively asking the same choice questions as a lottery-choice or dictator game setting would, however without direct compensation), or to capture "qualitative" self-assessment measures (such as "on a scale of 1 to 7, how willing to take risks are you?"). Regardless of the measure, "survey" techniques are distinguished by their lack of direct compensation to actions a subject takes.

By disconnecting compensation and choice, survey methods' theoretical strength and weakness is one in the same. On one hand, the lack of incentive gives the experimenter no inherent reason to believe that observed choices are, in fact, "true." In fact, if a subject's objective is to maximize their monetary payoff and minimize their effort (or time) spent, their payoff might be optimized by making choices arbitrarily, or by following a rule not directly related to their true preferences. At the same time, the lack of compensation means the theoretical concerns that plague currently used incentivized procedures are not relevant: there are no wealth, income, portfolio, or compound lottery effects, as no task impacts subject payment.

The appropriateness and validity of using a survey mechanism is itself likely strongly tied to what the data are intended to be used for. There is evidence that compensation can impact elicited preferences, which might be particularly important if the objective is to measure the existence of a paradox such as preference reversals, endowment effects, willingness to accept-willingness to pay gaps, probability weighting, and others.[108] At the same time, recent evidence supports the stability and effectiveness of survey methods for measuring basic utility parameters. Several papers, such as Dohmen, Falk, Huffman, Sunde, Schupp and Wagner (2011), Vieider et al. (2014), and Falk et al. (2016), compare hypothetical and incentivized responses from various populations in order to estimate predictive ability of hypothetical choice procedures, and find a significant correlation between the measures. These efforts have expanded to allow for large, cross-country comparisons of risk, time, ambiguity, and altruism preferences (most notable of which is the Global Preference Survey described in Falk et al. (2018)), enabling deeper research into the effects of language, culture, institutions, and other group-specific characteristics on these behavioral traits.[109]

3.6.2 Pay All Decisions

The most straightforward mechanism that assigns payments based on subject actions is to simply sum the outcomes of all decisions made. In other words, if an individual makes ten choices in ten different decision questions, their payment is simply their combined choices. Although there are some theoretical advantages of this mechanism, significant practical and theoretical limitations have limited its use in many situations.

The two key advantages of paying a subject based on every decision made are that it does not require the same behavioral assumption as alternatives (namely random selection mechanisms, described in the next section) and it is conceptually easier for certain subjects to understand. This former point, described at length in Azrieli et al. (2018), is key if the economist fears possible cross task contamination based on compound lottery effects. The

[108]See section 3.6.4 for more detail regarding the impact of compensation on elicited choice.

[109]The effectiveness of these methods has been subsequently tested by works such as Bauer et al. (2020), who find high validity for "quantitative" measures, but none for "qualitative" ones.

latter point has been used in experiments investigating experiential learning and choice feedback effects, such as Merlo and Schotter (2003). They point out the practical reality of most real-world decision making–that most decisions are small and yield immediate feedback–is better accounted for in a pay-all setting.

At the same time, there are significant theoretical and practical drawbacks that limit pay all designs' applicability more generally. As Azrieli and coauthors prove, pay all designs are sensitive to complementarities between decisions.[110] Moreover, paying multiple tasks can significantly increase the costs of conducting an experiment, often necessitating smaller payouts (which can, itself, impact subject incentives).

3.6.3 Random Task Selection

The present "best practices" mechanism's design is intended to address the income and wealth effect distortions that can be present in pay all experiments. What can (relatively interchangeably) be denoted a **random incentive system** (RIS), **random lottery incentive mechanism** (RLIM), and **pay one randomly** (POR) mechanism, the key feature of these experiments is that a subject is asked to perform multiple tasks, but is only compensated for a single randomly chosen action. This mechanism has been used to elicit preferences across a wide array of decision contexts and fields (see earlier sections for references). At the same time, it is necessary to differentiate these random selection mechanisms from stochastic termination procedures like those used in repeated game theory experiments, described in Chandrasekhar and Xandri (2017) and Deb et al. (2020), and in a decision environment as discussed in chapter II. To do so, I make the implicit assumption from the proofs in Azrieli et al. (2018) the delineating factor:

[110] As a straightforward, albeit highly stylized example, a subject's valuation of a left shoe is likely highly dependent on whether they have previously received the matching right shoe in an earlier round.

random incentive systems choose a task number for payment based on a pre-specified list of tasks.[111]

Effects of choosing one task at random on theoretical behavioral responses varies by context and the model being tested. Early works by Karni and Safra (1987) and Segal (1990), and later generalizations like Azrieli et al. (2018), recognized the importance of a version of the independence axiom for predicting how these mechanisms impact preferences. Informally, what matters for incentive compatibility is how individuals evaluate multi-stage lotteries. If subjects are non-expected utility maximizers and reduce compound lotteries (or at least use elements of the compound lottery in their subjective evaluation of the value of additional elements to the compound lottery), then random incentive systems can distort preferences through cross task contamination. If subjects satisfy either compound independence (from Segal (1990)) or statewise monotonicity (from Azrieli et al. (2018)), then random incentive systems are able to capture "true" underlying preferences. It should be noted that previous experiments have utilized this mechanism not based on a theoretical justification, but a practical one. For example, the lack of any credible alternative–combined with the practical impossibility associated with using one-task designs–was cited as primary reason for the POR's use in Harrison et al. (2017) (see footnote 9).

3.6.4 Evidence of Payment Procedure Effects

One underlying concern regarding experimental design is the potential of contamination, which, if present, can systematically (and unobservably) bias elicited preferences. With roots going back to the Wallis-Friedman critique of Thurstone, misalignments between theory, implementation, and subject incentives yield a continual potential threat to both internal and

[111]This list of tasks can, of course, be individual or treatment group-specific. The crucial factor is that the list of tasks is definable at the start of the experiment for each subject, something clearly impossible if the number of subgame repetitions or decision questions is not defined ahead of time.

external validity. Loomes (1999), while examining current and future directions of experimental and behavioral research designs, points to the underlying contradiction between certain methodologies and the theories they attempt to investigate.

Despite a number of large, well-designed laboratory experiments designed specifically to test for violations of either independence or reduction in decision making, their results have been decidedly mixed. Harrison, Martínez-Correa and Swarthout (2015) report evidence that reduction is violated when random incentive systems (in their case, a 1-in-40 design) is used–but no evidence of those same violations in the one-task treatment group's elicited preferences.

Other results identify suggestive evidence of the importance of compound independence (defined in Segal (1990) and elaborated on in Segal (1992)), and whether it or the reduction axiom is satisfied. Haering et al. (2020), for example, describes the impact of displaying reduced vs. compound lotteries on higher order risk preferences. More specifically, Haering et al. (2020)–drawing on the insights of Deck and Schlesinger (2016)–find that a sizable proportion of their subjects evaluate lotteries as "a combination of 'good' and 'bad' outcomes" and focus on particularly salient elements of them (such as the best and worst payoffs).

At the same time, another significant body of experiments has suggested procedures like random incentive systems are effective at capturing "true" preferences. One of the earliest, Starmer and Sugden (1991) compared subject choices across a random incentive system and one-real-one-hypothetical design, finding evidence of violations of reduction.[112]

These critiques are not limited to preferences elicited by lottery choices. In a wide-ranging meta-analysis of ultimatum and dictator game experiments, Engel (2011) finds a negative impact of repetition on giving (i.e. re-

[112] Starmer and Sugden (1991) is designed to follow up the methodological critique of Holt (1986) and argues that, by finding subjects violate reduction and not independence, the random incentive mechanism does not exhibit evidence of preference distortions.

peated dictator games induce senders to give less than one-shot games), however the impact of payment mechanism is less clear.[113] This broad takeaway has been echoed by subsequent work, such as Achtziger et al. (2015), who attribute a decrease in generosity in each round to an ego-depletion effect. Ben-Ner et al. (2008) compares donation amounts between hypothetical and incentivized dictator games and reports no significant difference in average donations.[114] This finding, however, contradicts both other experiments' findings (such as Bühren and Kundt (2015)), and the larger body of literature that observes a significant impact of stakes on dictator game decisions (see, for example, the meta analysis on this question by Larney et al. (2019).)

The dictator game environment is also potentially adversely impacted by contextual factors, such as identities of the recipient. In a one-shot dictator game, Eckel and Grossman (1996) compare average donations made by undergraduate students to another (anonymous) student and an established charity (a local affiliate of the American Red Cross), and find significantly higher average donations to the latter cause. This experimental setting has been repeated nearly 100 times, according to the meta analysis by Umer et al. (2022), who find the overall pattern holds: charitable causes receive larger shares of endowments than student recipients.[115]

3.7 Going Forward: New Experimental Methods

There have been a number of recent attempts to alter experimental structures. There are a wide variety of motivations behind this effort, including to more closely align the choice environment with a theory being tested, to capture behavioral aspects of decision-making, to measure alternative

[113] Significance depends on the method used to control for experiment-specific factors.

[114] Although this study's design raises questions regarding the validity of the comparison.

[115] Interestingly, Umer et al. (2022) makes no mention of whether experiments asked dictators to make one or multiple decisions—this is likely, at least partly, due to the lack of repeated-dictator-game experiments in the charitable-cause setting.

utility characteristics such as competitive preferences or attitudes towards cooperation, among many others. Here, I review several prominent deviations from the previously defined payment mechanisms intended to resolve problems of incentive compatibility. As in previous sections, I do not claim this is a comprehensive guide for all recent alternatives; instead, it is (I hope) a useful and instructive overview of some ways experimentalists have approached recent challenges. In each case, I briefly describe the alternative mechanism before mentioning some advantages and drawbacks.

3.7.1 PRINCE Mechanism

Chapter I defines and experimentally validates a new payment procedure for subjects making multiple decisions without imposing as restrictive as assumptions necessary in the mechanisms described above. Independently, another group has suggested a superficially similar procedure. Johnson et al. (2021) lays out what is called the "PRINCE" mechanism, which, like the Random Stopping Procedure detailed in section 1.3, attempts to develop an experimental environment of choice compensation that induces isolation between tasks. By "isolating" each decision, fewer assumptions must be imposed ex ante on subject preferences, which allows for more general tests of behavior.

As described in Johnson et al. (2021), the central change in PRINCE compared to current procedures is that the decision selection is determined at the start of the experiment. In effect, every possible choice that might be "real" is written on a slip of paper and placed in a separate envelope. The participant chooses one of these envelopes before answering any questions, and thus should not consider a future lottery over tasks.[116] Several experiments are then conducted with the PRINCE mechanism, intended to repli-

[116]The authors also describe a couple of other implementation aspects of the PRINCE mechanism that are intended to improve its function in practice, such as phrasing choices as "instructions to the experimenter" instead of abstract decisions. I focus on the predetermination not to minimize these other elements, but rather to highlight a potential concern.

cate prior experiments that found violations of rationality (such as preference reversals) or other "paradoxes" of choice (such as the endowment effect). Johnson and coauthors, using PRINCE, argue that many of these previously described deviations between rational behavior and observed choices are a result of procedural variance, and are more likely artifacts instead of 'true' paradoxes. Follow up work, such as Baillon et al. (2022) have demonstrated the theoretical advantages of resolving the "round-choice" lottery before the experiment begins in the more focused domain of analyzing ambiguity preferences.

While the replication exercises conducted by Johnson et al. (2021) are impressive and praiseworthy, there are theoretical and practical drawbacks to PRINCE's application that should be considered. The first, and most straightforward, is highlighted in a caveat included in Baillon et al. (2022): this approach requires subjects to view this pre-randomization as decisive. While it is quite possible that this is the case (and Johnson et al. (2021) suggest it should be more likely true than assuming independence outright), there is no theoretical grounding in decision models to definitely say one way or the other. This is a crucial assumption–and something that should be experimentally verified–as it undergirds all theoretical advantages over a standard random incentive mechanism.

Beyond this preference restriction, there are behavioral assumptions that are not clearly addressed by performing the randomization before questions are asked to subjects. Particularly relevant in settings like altruistic decisions, a body of literature has identified task interdependencies such as moral licensing, dissonance avoidance, warm glow, and more. In all of these situations, the possibility of other decisions being relevant induces subjects to behave systematically differently.

Finally, from an empirical perspective, it isn't immediately clear from the one direct test that PRINCE yields results closer to one task results than other common mechanisms. In their "Experiment 4," Johnson and coauthors replicate the design of Cox et al. (2015) with both a PRINCE group and random incentive system treatment, and do find PRINCE to yield results closer to Cox et al. (2015)'s one task results than their RIS treatment. At the same time, while instructive, are far from conclusive. Beyond the statistical non-significance of the results[117] differences between the pay one randomly treatments of Cox et al. (2015) and the RIS group in Johnson et al. (2021) make direct comparisons difficult to draw.

3.7.2 Pay Some Participants

An alternative to choosing one task to pay with certainty is choosing a random subset of subjects to receive compensation. There are a couple notable cases of this approach being taken. For example, Andersen et al. (2008) ask subjects to make four distinct multiple-price-list-style decisions, however only directly compensate 10% of their subject pool. Coffman (2016) asks subjects to complete a five minute survey, which yields a 1-in-75 chance of being drawn to win an $80 prize, which is divided between the student and a charity.[118]

The empirical effects of paying a subset of subjects is not as well researched as several of the other mechanisms described previously The one major exception is March et al. (2016), who do find support for paying only a subset of participants, if potential rewards are increased accordingly. However, for contexts like Coffman (2016) where the primary hypothesis involves the effects of changing subject information structure and not measuring unbiased preferences for donations directly, there are unlikely to be real drawbacks.

[117]Part of this is due to the small sample sizes used for both the RIS (25 subjects) and PRINCE (26 subjects) in both treatments.

[118]Coffman's primary interest is not in the amount of student donations per se, but rather how intermediation impacts sensitivity to charity quality in donations.

3.7.3 *Accumulative Best Choice*

Another recently proposed alternative to address the experimental design-incentive compatibility problem is the **accumulative best choice** (ABC) mechanism from Li (2016). Paraphrasing a bit, the major innovation in Li (2016) is to create a "tournament style" experiment, where the chosen option in one round is "dragged" to the next as a selectable option. From a motivational standpoint, the random stopping procedure defined in section 1.3 and the ABC mechanism are quite similar. Li avoids the "standard" incentive compatibility trap by relaxing the assumption of a fixed set of questions every subject observes, which is often considered exogenous. By allowing subjects to always choose their best possible option, which can be guaranteed to be a subject's compensation if they continue to select it in every round, the ABC mechanism avoids both income and compound lottery effects (since there is no direct compound lottery generated by the procedure).

As Li (2016) describes, the ABC is incentive compatible for measuring risk preferences under a range of non-expected utility theories. One drawback from this approach–aligning incentive compatibility with many current models of utility–is that it does not directly address the underlying problem of decision interdependence more generally. (In fact, the ABC makes subject payment directly and irrevocably intertwined with previous decision answers.) By not inducing separation between tasks, the ABC is not incentive compatible when testing all theories of decision making, potentially limiting its applicability to alternative notions of utility. For example, the ABC can still cause preference distortions for loss-aversion or reference dependent utility theories. In addition, the only decision an experimenter can predetermine is the first one (since all subsequent decisions include the choice from the previous round). Depending on the hypothesis being tested, this too can potentially limit the ABC's applicability for use in the lab.

4

Bibliography

Abdulkadiroglu, A., Che, Y.-K. and Yasuda, Y. (2015), 'Expanding "choice" in school choice', *American Economic Journal: Microeconomics* **7**(1), 1–42.
URL: *http://www.jstor.org/stable/24467033*

Abdulkadiroglu, A., Pathak, P. A. and Roth, A. E. (2009), 'Strategy-Proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match', *American Economic Review* **99**(5), 1954–1978.
URL: *https://ideas.repec.org/a/aea/aecrev/v99y2009i5p1954-78.html*

Abdulkadiroglu, A. and Sönmez, T. (2003), 'School choice: A mechanism design approach', *American Economic Review* **93**(3), 729–747.
URL: *http://www.aeaweb.org/articles?id=10.1257/000282803322157061*

Abdulkadiroglu, A. and Sönmez, T. (2013), *Matching Markets: Theory and Practice*, Vol. 1 of *Econometric Society Monographs*, Cambridge University Press, p. 3–47.

Abdulkadiroğlu, A., Pathak, P. A., Schellenberg, J. and Walters, C. R. (2020), 'Do parents value school effectiveness?', *American Economic Review* **110**(5), 1502–39.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.20172040*

Achtziger, A., Alós-Ferrer, C. and Wagner, A. K. (2015), 'Money, depletion, and prosociality in the dictator game.', *Journal of Neuroscience, Psychology, and Economics* **8**(1), 1–14.
URL: *https://doi.org/10.1037/npe0000031*

Ahn, D., Choi, S., Gale, D. and Kariv, S. (2014), 'Estimating ambiguity aversion in a portfolio choice experiment', *Quantitative Economics* **5**(2), 195–223.
URL: *https://doi.org/10.3982/qe243*

Aksoy, S., Adam Azzam, A., Coppersmith, C., Glass, J., Karaali, G., Zhao, X. and Zhu, X. (2013), 'School choice as a one-sided matching problem: Cardinal utilities and optimization', *arXiv e-prints* .
URL: *https://arxiv.org/abs/1304.7413*

Aksoy, S., Azzam, A., Coppersmith, C., Glass, J., Karaali, G., Zhao, X. and Zhu, X. (2015), 'Coalitions and cliques in the school choice problem', *In-*

volve, a Journal of Mathematics **8**(5), 801–823.
URL: *http://dx.doi.org/10.2140/involve.2015.8.801*

Allcott, H. (2011), 'Social norms and energy conservation', *Journal of Public Economics* **95**(9-10), 1082–1095.
URL: *https://doi.org/10.1016/j.jpubeco.2011.03.003*

Allcott, H. and Kessler, J. B. (2019), 'The welfare effects of nudges: A case study of energy use social comparisons', *American Economic Journal: Applied Economics* **11**(1), 236–276.
URL: *https://doi.org/10.1257/app.20170328*

Andersen, S., Harrison, G. W., Lau, M. I. and Rutström, E. E. (2008), 'Eliciting risk and time preferences', *Econometrica* **76**(3), 583–618.
URL: *http://www.jstor.org/stable/40056458*

Andersen, S., Harrison, G. W., Lau, M. I. and Rutström, E. E. (2006), 'Elicitation using multiple price list formats', *Experimental Economics* **9**(4), 383–405.
URL: *https://doi.org/10.1007/s10683-006-7055-6*

Anderson, L. R. and Mellor, J. M. (2008), 'Predicting health behaviors with an experimental measure of risk preference', *Journal of Health Economics* **27**(5), 1260–1274.
URL: *https://doi.org/10.1016/j.jhealeco.2008.05.011*

Andreoni, J. (1989), 'Giving with impure altruism: Applications to charity and ricardian equivalence', *Journal of Political Economy* **97**(6), 1447–1458.
URL: *http://www.jstor.org/stable/1833247*

Andreoni, J. and Bernheim, B. D. (2009), 'Social image and the 50-50 norm: A theoretical and experimental analysis of audience effects', *Econometrica* **77**(5), 1607–1636.
URL: *http://www.jstor.org/stable/25621371*

Andreoni, J., Harbaugh, W. T. and Vesterlund, L. (2010), Altruism in experiments, *in* 'Behavioural and Experimental Economics', Palgrave Macmillan UK, pp. 6–13.
URL: *https://doi.org/10.1057/9780230280786_2*

Andreoni, J., Rao, J. M. and Trachtman, H. (2017), 'Avoiding the ask: A field experiment on altruism, empathy, and charitable giving', *Journal of*

Political Economy **125**(3), 625–653.
URL: *https://doi.org/10.1086/691703*

Aridan, N., Malecek, N. J., Poldrack, R. A. and Schonberg, T. (2019), 'Neural correlates of effort-based valuation with prospective choices', *NeuroImage* **185**, 446 – 454.
URL: *http://www.sciencedirect.com/science/article/pii/S1053811918320238*

Armantier, O. and Treich, N. (2013), 'Eliciting beliefs: Proper scoring rules, incentives, stakes and hedging', *European Economic Review* **62**, 17–40.
URL: *https://doi.org/10.1016/j.euroecorev.2013.03.008*

Atkinson, A. C. (2015), Optimal experimental design, *in* 'International Encyclopedia of the Social & Behavioral Sciences', Elsevier, pp. 256–262.
URL: *https://doi.org/10.1016/b978-0-08-097086-8.42041-6*

Aygün, O. and Sönmez, T. (2012), Matching with Contracts: The Critical Role of Irrelevance of Rejected Contracts, Boston College Working Papers in Economics 804, Boston College Department of Economics.
URL: *https://ideas.repec.org/p/boc/bocoec/804.html*

Ayres, I. and Siegelman, P. (1995), 'Race and gender discrimination in bargaining for a new car', *The American Economic Review* **85**(3), 304–321.
URL: *http://www.jstor.org/stable/2118176*

Azrieli, Y., Chambers, C. and Healy, P. J. (2018), 'Incentives in experiments: A theoretical analysis', *Journal of Political Economy* **126**(4), 1472 – 1503.
URL: *https://EconPapers.repec.org/RePEc:ucp:jpolec:doi:10.1086/698136*

Azrieli, Y., Chambers, C. P. and Healy, P. J. (2019), 'Incentives in experiments with objective lotteries', *Experimental Economics* .
URL: *https://doi.org/10.1007/s10683-019-09607-0*

Babcock, B. A. (2015), 'Using cumulative prospect theory to explain anomalous crop insurance coverage choice', *American Journal of Agricultural Economics* **97**(5), 1371–1384.
URL: *https://doi.org/10.1093/ajae/aav032*

Baillon, A., Halevy, Y. and Li, C. (2022), 'Experimental elicitation of ambiguity attitude using the random incentive system', *Experimental Economics* .
URL: *https://doi.org/10.1007/s10683-021-09739-2*

Baillon, A., Huang, Z., Selim, A. and Wakker, P. P. (2018), 'Measuring ambi-

guity attitudes for all (natural) events', *Econometrica* **86**(5), 1839–1858.
URL: *http://www.jstor.org/stable/44955260*

Balinski, M. and Sönmez, T. (1999), 'A tale of two mechanisms: Student placement', *Journal of Economic Theory* **84**(1), 73–94.
URL: *https://EconPapers.repec.org/RePEc:eee:jetheo:v:84:y:1999:i:1:p:73-94*

Bando, K. (2012), 'Many-to-one matching markets with externalities among firms', *Journal of Mathematical Economics* **48**(1), 14–20.
URL: *https://EconPapers.repec.org/RePEc:eee:mateco:v:48:y:2012:i:1:p:14-20*

Barberis, N. C. (2013), 'Thirty years of prospect theory in economics: A review and assessment', *Journal of Economic Perspectives* **27**(1), 173–196.
URL: *https://doi.org/10.1257/jep.27.1.173*

Bardsley, N. (2007), 'Dictator game giving: altruism or artefact?', *Experimental Economics* **11**(2), 122–133.
URL: *https://doi.org/10.1007/s10683-007-9172-2*

Barham, B. L., Chavas, J.-P., Fitz, D., Salas, V. R. and Schechter, L. (2014), 'The roles of risk and ambiguity in technology adoption', *Journal of Economic Behavior & Organization* **97**, 204–218.
URL: *https://doi.org/10.1016/j.jebo.2013.06.014*

Barseghyan, L., Molinari, F., O'Donoghue, T. and Teitelbaum, J. C. (2018), 'Estimating risk preferences in the field', *Journal of Economic Literature* **56**(2), 501–564.
URL: *https://www.jstor.org/stable/26494195*

Bauer, M., Chytilová, J. and Miguel, E. (2020), 'Using survey questions to measure preferences: Lessons from an experimental validation in kenya', *European Economic Review* **127**, 103493.
URL: *https://doi.org/10.1016/j.euroecorev.2020.103493*

Ben-Ner, A., Kramer, A. and Levy, O. (2008), 'Economic and hypothetical dictator game experiments: Incentive effects at the individual level', *The Journal of Socio-Economics* **37**(5), 1775–1784.
URL: *https://doi.org/10.1016/j.socec.2007.11.004*

Bénabou, R., Falk, A., Henkel, L. and Tirole, J. (2020), Eliciting moral preferences: Theory and experiment, Technical report.

Benz, M. and Meier, S. (2008), 'Do people behave in experiments as in the

field? evidence from donations', *Experimental Economics* **11**(3), 268–281.
URL: *https://doi.org/10.1007/s10683-007-9192-y*

Bertrand, M. and Duflo, E. (2016), Field experiments on discrimination, Technical report.
URL: *https://doi.org/10.3386/w22014*

Bertrand, M. and Mullainathan, S. (2004), 'Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination', *American Economic Review* **94**(4), 991–1013.
URL: *https://doi.org/10.1257/0002828042002561*

Binmore, K. (1999), 'Why experiment in economics?', *The Economic Journal* **109**(453).
URL: *https://www.jstor.org/stable/2565582*

Blum, Y., Roth, A. E. and Rothblum, U. G. (1997), 'Vacancy chains and equilibration in senior-level labor markets', **76**(2), 362–411.
URL: *https://doi.org/10.1006/jeth.1997.2307*

Board, C. A. R., Jenkins, P. L., Phillips, T. J. and Waldman, J. (2004), *Califnornia Air Resources Board Assessment*, "California Air Resources Board".
URL: *https://www.arb.ca.gov/research/apr/reports/l3006.pdf*

Bocquého, G., Jacquet, F. and Reynaud, A. (2013), 'Expected utility or prospect theory maximisers? assessing farmers' risk behaviour from field-experiment data', *European Review of Agricultural Economics* **41**(1), 135–172.
URL: *https://doi.org/10.1093/erae/jbt006*

Bohm, P., Lindén, J. and Sonnegård, J. (2012), 'Eliciting reservation prices: Becker–degroot–marschak mechanisms vs. markets*', *The Economic Journal* **107**(443), 1079–1089.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0297.1997.tb00008.x*

Boksem, M. A. S., Meijman, T. F. and Lorist, M. M. (2006), 'Mental fatigue, motivation and action monitoring.', *Biological psychology* **72**, 123–32.

Bolton, G. and Ockenfels, A. (2012), 'Behavioral economic engineering', *Journal of Economic Psychology* **33**(3), 665–676.
URL: *https://EconPapers.repec.org/RePEc:eee:joepsy:v:33:y:2012:i:3:p:665-676*

Bönte, W., Lombardo, S. and Urbig, D. (2017), 'Economics meets psychol-

ogy: Experimental and self-reported measures of individual competitiveness', *Personality and Individual Differences* **116**, 179–185.
URL: *https://doi.org/10.1016/j.paid.2017.04.036*

Booij, A. S., van Praag, B. M. S. and van de Kuilen, G. (2009), 'A parametric analysis of prospect theory's functionals for the general population', *Theory and Decision* **68**(1-2), 115–148.
URL: *https://doi.org/10.1007/s11238-009-9144-4*

Bougherara, D., Gassmann, X., Piet, L. and Reynaud, A. (2017), 'Structural estimation of farmers' risk and ambiguity preferences: a field experiment', *European Review of Agricultural Economics* **44**(5), 782–808.
URL: *https://doi.org/10.1093/erae/jbx011*

Braden, J. B. and Kolstad, C. D. (1991), *Measuring the demand for environmental quality / edited by John B. Braden, Charles D. Kolstad*, North-Holland ; Distributors for the U.S. and Canada, Elsevier Science Pub. Co Amsterdam ; New York : New York, N.Y., U.S.A.

Brown, A. L. and Healy, P. J. (2018), 'Separated decisions', *European Economic Review* **101**(C), 20–34.
URL: *https://ideas.repec.org/a/eee/eecrev/v101y2018icp20-34.html*

Brunette, M. and Ngouhouo-Poufoun, J. (2019), Are risk preferences stable ? A field experiment in Congo Basin countries, Technical report.

Budish, E. and Cantillon, E. (2012), 'The multi-unit assignment problem: Theory and evidence from course allocation at harvard', *American Economic Review* **102**(5), 2237–71.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.102.5.2237*

Bühren, C. and Kundt, T. C. (2015), 'Imagine being a nice guy: A note on hypothetical vs. incentivized social preferences', *Judgment and Decision Making* **10**(2), 185–190.
URL: *https://ideas.repec.org/a/jdm/journl/v10y2015i2p185-190.html*

Bühren, C., Meier, F. and Pleßner, M. (2021), 'Ambiguity aversion: bibliometric analysis and literature review of the last 60 years', *Management Review Quarterly* .
URL: *https://doi.org/10.1007/s11301-021-00250-9*

Burks, S. V., Carpenter, J. P., Goette, L. and Rustichini, A. (2009), 'Cognitive skills affect economic preferences, strategic behavior, and job attachment',

Proceedings of the National Academy of Sciences **106**(19), 7745–7750.
URL: *https://doi.org/10.1073/pnas.0812360106*

Camerer, C. F., Ho, T.-H. and Chong, J.-K. (2004), 'A cognitive hierarchy model of games', *The Quarterly Journal of Economics* **119**(3), 861–898.
URL: *https://doi.org/10.1162/0033553041502225*

Camerer, C. F. and Hogarth, R. M. (1999), 'The effects of financial incentives in experiments: A review and capital-labor-production framework', *Journal of Risk and Uncertainty* **19**(1/3), 7–42.
URL: *https://doi.org/10.1023/a:1007850605129*

Camerer, C. and Thaler, R. H. (1995), 'Anomalies: Ultimatums, dictators and manners', *Journal of Economic Perspectives* **9**(2), 209–219.
URL: *https://doi.org/10.1257/jep.9.2.209*

Caplin, A. and Dean, M. (2015), 'Revealed preference, rational inattention, and costly information acquisition', *American Economic Review* **105**(7), 2183–2203.
URL: *https://doi.org/10.1257/aer.20140117*

Card, D. and Krueger, A. B. (2000), 'Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania: Reply', *The American Economic Review* **90**(5), 1397–1420.
URL: *http://www.jstor.org/stable/2677856*

Carlson, R. W., Aknin, L. B. and Liotti, M. (2015), 'When is giving an impulse? an ERP investigation of intuitive prosocial behavior', **11**(7), 1121–1129.
URL: *https://doi.org/10.1093/scan/nsv077*

Carpenter, J. P. and Huet-Vaughn, E. (2019), 'Real-effort tasks', *Handbook of Research Methods and Applications in Experimental Economics* .

Chakravarty, S., Friedman, D., Gupta, G., Hatekar, N., Mitra, S. and Sunder, S. (2011), '1 emergence of experimental economics', *Economic and Political Weekly* **46**(35), 41–46.
URL: *http://www.jstor.org/stable/23017906*

Chandrasekhar, A. G. and Xandri, J. P. (2017), 'A note on payments in the lab for infinite horizon dynamic games with discounting'.
URL: *https://scholar.princeton.edu/jxandri/publications/note-payments-lab-infinite-horizon-dynamic-games-discounting*

Charlotte Mecklenburg Board of Education and CMS (2011), *Valuing Learning Environments: Providing Sufficient Capacity in Facilities and Resources, Capital Needs Assessment*, Charlotte Mecklenburg Schools.

Charness, G., Garcia, T., Offerman, T. and Villeval, M. C. (2020), 'Do measures of risk attitude in the laboratory predict behavior under risk in and outside of the laboratory?', *Journal of Risk and Uncertainty* **60**(2), 99–123.
URL: *https://doi.org/10.1007/s11166-020-09325-6*

Charness, G., Gneezy, U. and Halladay, B. (2016), 'Experimental methods: Pay one or pay all', *Journal of Economic Behavior & Organization* **131**(PA), 141–150.
URL: *https://EconPapers.repec.org/RePEc:eee:jeborg:v:131:y:2016:i:pa:p:141-150*

Charness, G., Gneezy, U. and Henderson, A. (2018), 'Experimental methods: Measuring effort in economics experiments', *Journal of Economic Behavior & Organization* **149**, 74–87.
URL: *https://doi.org/10.1016/j.jebo.2018.02.024*

Charness, G., Gneezy, U. and Imas, A. (2013), 'Experimental methods: Eliciting risk preferences', *Journal of Economic Behavior & Organization* **87**, 43–51.
URL: *https://doi.org/10.1016/j.jebo.2012.12.023*

Charness, G., Gneezy, U. and Kuhn, M. A. (2013), 'Experimental methods: Extra-laboratory experiments-extending the reach of experimental economics', *Journal of Economic Behavior & Organization* **91**, 93–100.
URL: *https://doi.org/10.1016/j.jebo.2013.04.002*

Charness, G. and Rabin, M. (2002), 'Understanding social preferences with simple tests', *The Quarterly Journal of Economics* **117**(3), 817–869.
URL: *http://www.jstor.org/stable/4132490*

Charness, G. and Viceisza, A. (2016), 'Three Risk-elicitation Methods in the Field - Evidence from Rural Senegal', *Review of Behavioral Economics* **3**(2), 145–171.
URL: *https://ideas.repec.org/a/now/jnlrbe/105.00000046.html*

Charness, G. and Villeval, M.-C. (2009), 'Cooperation and competition in intergenerational experiments in the field and the laboratory', *American*

Economic Review **99**(3), 956–978.
URL: *https://doi.org/10.1257/aer.99.3.956*

Chen, N. and Li, M. (2013), Ties matter: improving efficiency in course allocation by introducing ties, MPRA Paper 47031, University Library of Munich, Germany.
URL: *https://ideas.repec.org/p/pra/mprapa/47031.html*

Chew, S. H., Miao, B. and Zhong, S. (2017), 'Partial ambiguity', *Econometrica* **85**(4), 1239–1260.
URL: *https://doi.org/10.3982/ecta13239*

Choi, S., Fisman, R., Gale, D. and Kariv, S. (2007), 'Consistency and heterogeneity of individual behavior under uncertainty', *American Economic Review* **97**(5), 1921–1938.
URL: *https://doi.org/10.1257/aer.97.5.1921*

Clot, S., Grolleau, G. and Ibanez, L. (2013), 'Self-Licensing and Financial Rewards: Is Morality For Sale?', *Economics Bulletin* **33**(3), 2298–2306.
URL: *https://ideas.repec.org/a/ebl/ecbull/eb-13-00385.html*

Coffman, K. B., Coffman, L. C. and Ericson, K. M. M. (2017), 'The size of the LGBT population and the magnitude of antigay sentiment are substantially underestimated', *Management Science* **63**(10), 3168–3186.
URL: *https://doi.org/10.1287/mnsc.2016.2503*

Coffman, L. C. (2016), 'Fundraising Intermediaries Inhibit Quality-Driven Charitable Donations', *Economic Inquiry* **55**(1), 409–424.
URL: *https://doi.org/10.1111/ecin.12379*

Coffman, L. C. (2019), 'Expectations do not affect punishment', *Journal of the Economic Science Association* **5**(2), 182–196.
URL: *https://doi.org/10.1007/s40881-019-00079-9*

Coffman, L. C., Conlon, J. J., Featherstone, C. R. and Kessler, J. B. (2019), 'Liquidity affects job choice: Evidence from teach for america', *The Quarterly Journal of Economics* **134**(4), 2203–2236.
URL: *https://doi.org/10.1093/qje/qjz018*

Cohen, J., Ericson, K. M., Laibson, D. and White, J. M. (2020), 'Measuring time preferences', *Journal of Economic Literature* **58**(2), 299–347.
URL: *https://doi.org/10.1257/jel.20191074*

Cook, J. H., Chatterjee, S., Sur, D. and Whittington, D. (2011), 'Measuring

risk aversion among the urban poor in kolkata, india', *SSRN Electronic Journal* .
URL: *https://doi.org/10.2139/ssrn.1956178*

Coombs, C., Bezembinder, T. and Goode, F. (1967), 'Testing expectation theories of decision making without measuring utility or subjective probability', *Journal of Mathematical Psychology* **4**(1), 72–103.
URL: *https://doi.org/10.1016/0022-2496(67)90042-9*

Cooper, D. and Kagel, J. (2016), 'Other regarding preferences: A selective survey of experimental results', *The handbook of experimental economics* **2**.

Cox, J. C., Sadiraj, V. and Schmidt, U. (2015), 'Paradoxes and mechanisms for choice under risk', *Experimental Economics* **18**(2), 215–250.
URL: *https://doi.org/10.1007/s10683-014-9398-8*

Cox, J. C., Sadiraj, V., Vogt, B. and Dasgupta, U. (2012), 'Is there a plausible theory for decision under risk? a dual calibration critique', *Economic Theory* **54**(2), 305–333.
URL: *https://doi.org/10.1007/s00199-012-0712-4*

Cox, J., Sadiraj, V. and Schmidt, U. (2013), Paradoxes and mechanisms for choice under risk, Experimental Economics Center Working Paper Series 2011-07, Experimental Economics Center, Andrew Young School of Policy Studies, Georgia State University.
URL: *https://EconPapers.repec.org/RePEc:exc:wpaper:2011-07*

Cox, J., Sadiraj, V. and Schmidt, U. (2014), Asymmetrically dominated choice problems, the isolation hypothesis and random incentive mechanisms, Mpra paper, University Library of Munich, Germany.
URL: *https://EconPapers.repec.org/RePEc:pra:mprapa:54722*

Cramton, P. (1998), The fcc spectrum auctions: An early assessment, Papers of peter cramton, University of Maryland, Department of Economics - Peter Cramton.
URL: *https://EconPapers.repec.org/RePEc:pcc:pccumd:97jemsfcc*

Crawford, V. P. (1997), Theory and experiment in the analysis of strategic interaction, *in* D. M. Kreps and K. F. Wallis, eds, 'Advances in Economics and Econometrics: Theory and Applications', Cambridge University Press, pp. 206–242.
URL: *https://doi.org/10.1017/ccol521580110.007*

Croson, R. (2002), 'Why and how to experiment: Methodologies from experimental economics', *University of Illinois Law Review* **2002**, 921–945.

Crumpler, H. and Grossman, P. J. (2008), 'An experimental test of warm glow giving', *Journal of Public Economics* **92**(5-6), 1011–1021.
URL: *https://doi.org/10.1016/j.jpubeco.2007.12.014*

Cubitt, R., van de Kuilen, G. and Mukerji, S. (2018), 'The strength of sensitivity to ambiguity', *Theory and Decision* **85**(3-4), 275–302.
URL: *https://doi.org/10.1007/s11238-018-9657-9*

Dana, J., Cain, D. M. and Dawes, R. M. (2006*a*), 'What you don't know won't hurt me: Costly (but quiet) exit in dictator games', *Organizational Behavior and Human Decision Processes* **100**(2), 193–201.
URL: *https://EconPapers.repec.org/RePEc:eee:jobhdp:v:100:y:2006:i:2:p:193-201*

Dana, J., Cain, D. M. and Dawes, R. M. (2006*b*), 'What you don't know won't hurt me: Costly (but quiet) exit in dictator games', *Organizational Behavior and Human Decision Processes* **100**(2), 193–201.
URL: *https://doi.org/10.1016/j.obhdp.2005.10.001*

Davis, D. D. and Holt, C. A. (1992), *Experimental Economics*, hardcover edn, Princeton University Press.

de Bekker-Grob, E. W., Ryan, M. and Gerard, K. (2010), 'Discrete choice experiments in health economics: a review of the literature', *Health Economics* **21**(2), 145–172.
URL: *https://doi.org/10.1002/hec.1697*

Dean, M. and McNeil, J. M. (2014), Preference for flexibility and random choice: an experimental analysis, Technical report.

Dean, M. and Ortoleva, P. (2019), 'The empirical relationship between nonstandard economic behaviors', *Proceedings of the National Academy of Sciences* **116**(33), 16262–16267.
URL: *https://doi.org/10.1073/pnas.1821353116*

Deb, J., Sugaya, T. and Wolitzky, A. (2020), 'The folk theorem in repeated games with anonymous random matching', *Econometrica* **88**(3), 917–964.
URL: *https://doi.org/10.3982/ecta16680*

Deck, C. and Jahedi, S. (2015), 'The effect of cognitive load on economic decision making: A survey and new experiments', *European Economic Re-*

view **78**, 97–119.
URL: *https://doi.org/10.1016/j.euroecorev.2015.05.004*

Deck, C. and Schlesinger, H. (2016), On the Robustness of Higher Order Risk Preferences, Technical report.

Dhingra, N., Gorn, Z., Kener, A. and Dana, J. (2012), 'The default pull: An experimental demonstration of subtle default effects on preferences', *Judgment and Decision Making* **7**.

Dillenberger, D. and Segal, U. (2017), 'Skewed noise', *Journal of Economic Theory* **169**, 344–364.
URL: *https://doi.org/10.1016/j.jet.2017.02.005*

Dimmock, S. G., Kouwenberg, R. and Wakker, P. P. (2016), 'Ambiguity attitudes in a large representative sample', *Management Science* **62**(5), 1363–1380.
URL: *https://doi.org/10.1287/mnsc.2015.2198*

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J. and Wagner, G. G. (2011), 'Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences, journal = Journal of the European Economic Association', **9**(3), 522–550.
URL: *https://doi.org/10.1111/j.1542-4774.2011.01015.x*

Dohmen, T., Huffman, D., Schupp, J., Falk, A., Sunde, U. and Wagner, G. G. (2011), 'Individual risk attitudes: Measurement, determinants, and behavioral consequences', *Journal of the European Economic Association* **9**(3), 522–550.
URL: *http://www.jstor.org/stable/25836078*

Dreber, A., Fudenberg, D. and Rand, D. G. (2014), 'Who cooperates in repeated games: The role of altruism, inequity aversion, and demographics', *Journal of Economic Behavior & Organization* **98**, 41–55.
URL: *https://doi.org/10.1016/j.jebo.2013.12.007*

Duflo, E. (2001), 'Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment', *American Economic Review* **91**(4), 795–813.
URL: *https://doi.org/10.1257/aer.91.4.795*

Dur, U. (2012), A Characterization of the Top Trading Cycles Mechanism for the School Choice Problem, MPRA Paper 41366, University Library of

Munich, Germany.
URL: *https://ideas.repec.org/p/pra/mprapa/41366.html*

Dur, U., Pathak, P. A. and Sönmez, T. (2016), Explicit vs. statistical preferential treatment in affirmative action: Theory and evidence from chicago's exam schools, Working Paper 22109, National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w22109*

Echenique, F. (2012), 'Contracts versus salaries in matching', **102**(1), 594–601.
URL: *https://doi.org/10.1257/aer.102.1.594*

Echenique, F., Imai, T. and Saito, K. (2019), Decision making under uncertainty: An experimental study in market settings, Technical report.
URL: *https://arxiv.org/abs/1911.00946*

Eckel, C. C. (2019), 'Measuring individual risk preferences', *IZA World of Labor* pp. 454–454.
URL: *https://ideas.repec.org/a/iza/izawol/journl2019n454.html*

Eckel, C. C. and Grossman, P. J. (1996), 'Altruism in anonymous dictator games', *Games and Economic Behavior* **16**(2), 181–191.
URL: *https://doi.org/10.1006/game.1996.0081*

Eckel, C., Priday, B. and Wilson, R. (2018), 'Charity begins at home: A lab-in-the-field experiment on charitable giving', *Games* **9**(4), 95.
URL: *https://doi.org/10.3390/g9040095*

Ehlers, L. and Morrill, T. (2017), (il)legal assignments in school choice, Cahiers de recherche, Universite de Montreal, Departement de sciences economiques.
URL: *https://EconPapers.repec.org/RePEc:mtl:montde:2017-02*

Elliott, S. R. (1998), 'Experiments in decision-making under risk and uncertainty: Thinking outside the box', *Managerial and Decision Economics* **19**(4/5), 239–257.
URL: *http://www.jstor.org/stable/3108237*

Ellsberg, D. (1961), 'Risk, ambiguity, and the savage axioms', *The Quarterly Journal of Economics* **75**(4), 643.
URL: *https://doi.org/10.2307/1884324*

Engel, C. (2011), 'Dictator games: a meta study', *Experimental Economics*

14(4), 583–610.
URL: *https://doi.org/10.1007/s10683-011-9283-7*

Epple, D., Jha, A. and Sieg, H. (2018), 'The superintendent's dilemma: Managing school district capacity as parents vote with their feet', *Quantitative Economics* **9**(1), 483–520.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.3982/QE592*

Erdil, A. and Ergin, H. (2008), 'What's the matter with tie-breaking? improving efficiency in school choice', *The American Economic Review* **98**(3), 669–689.
URL: *http://www.jstor.org/stable/29730091*

Ert, E. and Erev, I. (2011), 'On the descriptive value of loss aversion in decisions under risk', *SSRN Electronic Journal* .
URL: *https://doi.org/10.2139/ssrn.1012022*

European Conference on the History of Economics (5th : 1999 : Cachan, France) (2005), *The Experiment in the History of Economics*, Routledge.
URL: *https://doi.org/10.4324/9780203023594*

Exley, C. L. and Kessler, J. B. (2018), Equity concerns are narrowly framed, Working Paper 25326, National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w25326*

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D. and Sunde, U. (2018), 'Global evidence on economic preferences', *The Quarterly Journal of Economics* **133**(4), 1645–1692.
URL: *https://doi.org/10.1093/qje/qjy013*

Falk, A., Becker, A., Dohmen, T., Huffman, D. and Sunde, U. (2016), 'The preference survey module: A validated instrument for measuring risk, time, and social preferences', *SSRN Electronic Journal* .
URL: *https://doi.org/10.2139/ssrn.2725035*

Fallucchi, F., Nosenzo, D. and Reuben, E. (2021), Measuring preferences for competition with experimentally-validated survey questions, Working-Paper 2021-12, LISER, Luxembourg.

Featherstone, C. R. and Niederle, M. (2016), 'Boston versus deferred acceptance in an interim setting: An experimental investigation', *Games and Economic Behavior* **100**, 353 – 375.
URL: *http://www.sciencedirect.com/science/article/pii/S0899825616301208*

Felix, O.-G. and Reiner, E. (2008), 'Fairness in Extended Dictator Game Experiments', *The B.E. Journal of Economic Analysis & Policy* **8**(1), 1–21.
URL: *https://ideas.repec.org/a/bpj/bejeap/v8y2008i1n16.html*

Filippin, A. and Gioia, F. (2018), 'Competition and subsequent risk-taking behaviour: Heterogeneity across gender and outcomes', *Journal of Behavioral and Experimental Economics* **75**, 84–94.
URL: *https://doi.org/10.1016/j.socec.2018.05.003*

Fisher, I. (1892), Mathematical investigations in the theory of value and prices, Thesis, Yale University.
URL: *https://fraser.stlouisfed.org/files/docs/publications/books/mathematical_fisher.pdf*

Fisher, J. C. and Hafalir, I. E. (2016), 'Matching with aggregate externalities', *Mathematical Social Sciences* **81**(C), 1–7.
URL: *https://ideas.repec.org/a/eee/matsoc/v81y2016icp1-7.html*

Forsythe, R., Horowitz, J. L., Savin, N. and Sefton, M. (1994), 'Fairness in simple bargaining experiments', *Games and Economic Behavior* **6**(3), 347–369.
URL: *https://doi.org/10.1006/game.1994.1021*

Frechette, G. R., Sarnoff, K. and Yariv, L. (2021), Experimental economics: Past and future, Unpublished manuscript.
URL: *https://lyariv.mycpanel.princeton.edu/papers/ExperimentsPastFuture.pdf*

Fréchette, G. R. and Yuksel, S. (2017), 'Infinitely repeated games in the laboratory: four perspectives on discounting and random termination', *Experimental Economics* **20**(2), 279–308.
URL: *https://doi.org/10.1007/s10683-016-9494-z*

Freeman, D., Halevy, Y. and Kneeland, T. (2015), Eliciting Risk Preferences Using Choice Lists, Microeconomics.ca working papers yoram_halevy-2015-9, Vancouver School of Economics.
URL: *https://ideas.repec.org/p/ubc/pmicro/yoram_halevy-2015-9.html*

Freeman, D., Manzini, P., Mariotti, M. and Mittone, L. (2016), 'Procedures for eliciting time preferences', *Journal of Economic Behavior & Organization* **126**, 235–242.
URL: *https://doi.org/10.1016/j.jebo.2016.03.017*

Friesen, L. and Gangadharan, L. (2013), 'Environmental markets: what do

we learn from the lab?', *Journal of Economic Surveys* **27**(3), 515–535.
URL: *https://doi.org/10.1111/joes.12021*

Frohlich, N., Oppenheimer, J. and Moore, J. B. (2001), 'Some doubts about measuring self-interest using dictator experiments: the costs of anonymity', *Journal of Economic Behavior & Organization* **46**(3), 271–290.
URL: *https://doi.org/10.1016/s0167-2681(01)00178-0*

Gächter, S., Huang, L. and Sefton, M. (2015), 'Combining "real effort" with induced effort costs: the ball-catching task', *Experimental Economics* **19**(4), 687–712.
URL: *https://doi.org/10.1007/s10683-015-9465-9*

Gächter, S., Johnson, E. J. and Herrmann, A. (2021), 'Individual-level loss aversion in riskless and risky choices', *Theory and Decision* .
URL: *https://doi.org/10.1007/s11238-021-09839-8*

Galarza, F. (2009), Choices under Risk in Rural Peru, MPRA Paper 17708, University Library of Munich, Germany.
URL: *https://ideas.repec.org/p/pra/mprapa/17708.html*

Gale, D. and Shapley, L. S. (1962), 'College admissions and the stability of marriage', *The American Mathematical Monthly* **69**(1), 9–15.
URL: *http://www.jstor.org/stable/2312726*

Galizzi, M. M., Miraldo, M., Stavropoulou, C. and van der Pol, M. (2016), 'Doctor-patient differences in risk and time preferences: A field experiment', *Journal of Health Economics* **50**, 171–182.
URL: *https://doi.org/10.1016/j.jhealeco.2016.10.001*

García-Gallego, A., Georgantzis, N. and Ruiz-Martos, M. J. (2019), 'The heaven dictator game: Costless taking or giving', *Journal of Behavioral and Experimental Economics* **82**, 101449.
URL: *https://doi.org/10.1016/j.socec.2019.101449*

Geng, S. (2016), 'Decision Time, Consideration Time, And Status Quo Bias', *Economic Inquiry* **54**(1), 433–449.

Georgescu-Roegen, N. (1936), 'The pure theory of consumer's behaviour', *The Quarterly Journal of Economics* **50**(4), 545.
URL: *https://doi.org/10.2307/1891094*

Ghesla, C., Grieder, M. and Schmitz, J. (2019), 'Nudge for good? choice

defaults and spillover effects', *Frontiers in Psychology* **10**.
URL: *https://doi.org/10.3389/fpsyg.2019.00178*

Giamattei, M., Yahosseini, K. S., Gächter, S. and Molleman, L. (2020), 'LIONESS lab: a free web-based platform for conducting interactive experiments online', *Journal of the Economic Science Association* **6**(1), 95–111.
URL: *https://doi.org/10.1007/s40881-020-00087-0*

Gilboy, E. W. (1932), 'Demand curves by personal estimate', *The Quarterly Journal of Economics* **46**(2), 376–384.
URL: *http://www.jstor.org/stable/1883236*

Gneezy, U. and Imas, A. (2017), Lab in the field, *in* 'Handbook of Field Experiments', Elsevier, pp. 439–464.
URL: *https://doi.org/10.1016/bs.hefe.2016.08.003*

Gneezy, U., Leonard, K. L. and List, J. A. (2009), 'Gender differences in competition: Evidence from a matrilineal and a patriarchal society', *Econometrica* **77**(5), 1637–1664.
URL: *https://doi.org/10.3982/ecta6690*

Gneezy, U., Meier, S. and Rey-Biel, P. (2011), 'When and why incentives (don't) work to modify behavior', *The Journal of Economic Perspectives* **25**(4), 191–209.
URL: *http://www.jstor.org/stable/41337236*

Gneezy, U., Niederle, M. and Rustichini, A. (2003), 'Performance in competitive environments: Gender differences', *The Quarterly Journal of Economics* **118**(3), 1049–1074.
URL: *https://doi.org/10.1162/00335530360698496*

Gneezy, U. and Potters, J. (1997), 'An experiment on risk taking and evaluation periods', *The Quarterly Journal of Economics* **112**(2), 631–645.
URL: *http://www.jstor.org/stable/2951248*

Gneezy, U. and Rustichini, A. (2004), 'Gender and competition at a young age', *American Economic Review* **94**(2), 377–381.
URL: *https://doi.org/10.1257/0002828041301821*

Goldin, C. and Rouse, C. (2000), 'Orchestrating impartiality: The impact of 'blind' auditions on female musicians', *American Economic Review* **90**(4), 715–741.
URL: *https://doi.org/10.1257/aer.90.4.715*

Gonzalez, R. and Wu, G. (1999), 'On the shape of the probability weighting function', *Cognitive Psychology* **38**(1), 129–166.
URL: *https://doi.org/10.1006/cogp.1998.0710*

Grinstead, C. M. and Snell, J. L. (1998), *Introduction to probability*, 2 edn, American Mathematical Society.

Grossklags, J. (2007), Experimental economics and experimental computer science, *in* 'Proceedings of the 2007 workshop on Experimental computer science - ExpCS'07', ACM Press.
URL: *https://doi.org/10.1145/1281700.1281711*

GURIEV, S. (2001), 'On microfoundations of the dual theory of choice', *The Geneva Papers on Risk and Insurance Theory* **26**(2), 117–137.
URL: *http://www.jstor.org/stable/41953401*

Güth, W. and Kliemt, H. (2017), *Experimental Economics—A Philosophical Perspective*, Oxford University Press.
URL: *https://doi.org/10.1093/oxfordhb/9780199935314.013.16*

Haering, A., Heinrich, T. and Mayrhofer, T. (2020), 'Exploring the consistency of higher order risk preferences', *International Economic Review* **61**(1), 283–320.
URL: *https://doi.org/10.1111/iere.12424*

Hahn, R. and Metcalfe, R. (2016), 'The Impact of Behavioral Science Experiments on Energy Policy', *Economics of Energy & Environmental Policy* **0**(Number 2).
URL: *https://ideas.repec.org/a/aen/eeepjl/eeep5-2-hahn.html*

Harrison, G. and Swarthout, J. (2014), 'Experimental payment protocols and the bipolar behaviorist', *Theory and Decision* **77**(3), 423–438.
URL: *https://EconPapers.repec.org/RePEc:kap:theord:v:77:y:2014:i:3:p:423-438*

Harrison, G. W., Lau, M. I., Ross, D. and Swarthout, J. T. (2017), 'Small stakes risk aversion in the laboratory: A reconsideration', *Economics Letters* **160**, 24–28.
URL: *https://doi.org/10.1016/j.econlet.2017.08.003*

Harrison, G. W. and List, J. A. (2004), 'Field experiments', *Journal of Economic Literature* **42**(4), 1009–1055.
URL: *https://doi.org/10.1257/0022051043004577*

Harrison, G. W., List, J. A. and Towe, C. (2007), 'Naturally occurring prefer-

ences and exogenous laboratory experiments: A case study of risk aversion', *Econometrica* **75**(2), 433–458.
URL: *http://www.jstor.org/stable/4501996*

Harrison, G. W., Martínez-Correa, J. and Swarthout, J. T. (2015), 'Reduction of compound lotteries with objective probabilities: Theory and evidence', *Journal of Economic Behavior & Organization* **119**, 32–55.
URL: *https://doi.org/10.1016/j.jebo.2015.07.012*

Harrison, G. W., Martínez-Correa, J. and Swarthout, J. T. (2015), 'Reduction of compound lotteries with objective probabilities: Theory and evidence', *Journal of Economic Behavior & Organization* **119**(C), 32–55.
URL: *https://ideas.repec.org/a/eee/jeborg/v119y2015icp32-55.html*

Harrison, G. W. and Swarthout, J. T. (2016), Cumulative Prospect Theory in the Laboratory: A Reconsideration, Experimental Economics Center Working Paper Series 2016-04, Experimental Economics Center, Andrew Young School of Policy Studies, Georgia State University.
URL: *https://ideas.repec.org/p/exc/wpaper/2016-04.html*

Hashimoto, T., Hirata, D., Kesten, O., Kurino, M. and Ünver, M. U. (2014), 'Two axiomatic approaches to the probabilistic serial mechanism', *Theoretical Economics* **9**(1).
URL: *https://ideas.repec.org/a/the/publsh/1010.html*

Hatfield, J. W. and Kojima, F. (2008), 'Matching with contracts: Comment', **98**(3), 1189–1194.
URL: *https://doi.org/10.1257/aer.98.3.1189*

Hatfield, J. W. and Kojima, F. (2010), 'Substitutes and stability for matching with contracts', **145**(5), 1704–1723.
URL: *https://doi.org/10.1016/j.jet.2010.01.007*

Hatfield, J. W. and Milgrom, P. R. (2005), 'Matching with contracts', *American Economic Review* **95**(4), 913–935.
URL: *http://www.aeaweb.org/articles?id=10.1257/0002828054825466*

Hauber, A. B., González, J. M., Groothuis-Oudshoorn, C. G., Prior, T., Marshall, D. A., Cunningham, C., IJzerman, M. J. and Bridges, J. F. (2016), 'Statistical methods for the analysis of discrete choice experiments: A report of the ISPOR conjoint analysis good research practices task force',

Value in Health **19**(4), 300–315.
URL: *https://doi.org/10.1016/j.jval.2016.04.004*

He, T.-S. and Hong, F. (2018), 'Risk breeds risk aversion', *Experimental Economics* **21**(4), 815–835.
URL: *https://doi.org/10.1007/s10683-017-9553-0*

Hebert, B. and Woodford, M. (2019), Rational inattention when decisions take time, NBER Working Papers 26415, National Bureau of Economic Research, Inc.
URL: *https://EconPapers.repec.org/RePEc:nbr:nberwo:26415*

Hellerstein, D., Higgins, N. and Horowitz, J. (2013), 'The predictive power of risk preference measures for farming decisions', *European Review of Agricultural Economics* **40**(5), 807–833.
URL: *https://doi.org/10.1093/erae/jbs043*

Hertwig, R. and Ortmann, A. (2001), 'Experimental practices in economics: A methodological challenge for psychologists?', *Behavioral and Brain Sciences* **24**(3), 383–403.
URL: *https://doi.org/10.1017/s0140525x01004149*

Hey, J. D. and Lee, J. (2005), 'Do subjects separate (or are they sophisticated)?', *Experimental Economics* **8**(3), 233–265.
URL: *https://doi.org/10.1007/s10683-005-1465-8*

Hey, J. D. and Orme, C. (1994), 'Investigating generalizations of expected utility theory using experimental data', *Econometrica* **62**(6), 1291.
URL: *https://doi.org/10.2307/2951750*

Hoffman, E., McCabe, K. and Smith, V. L. (1996), 'Social distance and other-regarding behavior in dictator games', *The American Economic Review* **86**(3), 653–660.
URL: *http://www.jstor.org/stable/2118218*

Hoffman, E. and Spitzer, M. L. (1985), 'Experimental law and economics: An introduction', *Columbia Law Review* **85**(5), 991.
URL: *https://doi.org/10.2307/1122460*

Holt, C. A. (1986), 'Preference reversals and the independence axiom', *The American Economic Review* **76**(3), 508–515.
URL: *http://www.jstor.org/stable/1813367*

Holt, C. A. and Laury, S. K. (2002), 'Risk aversion and incentive effects',

American Economic Review **92**(5), 1644–1655.
URL: *https://doi.org/10.1257/000282802762024700*

Holt, C. A. and Laury, S. K. (2014), Assessment and estimation of risk preferences, *in* 'Handbook of the Economics of Risk and Uncertainty', Elsevier, pp. 135–201.
URL: *https://doi.org/10.1016/b978-0-444-53685-3.00004-0*

Holzmeister, F. and Stefan, M. (2020), 'The risk elicitation puzzle revisited: Across-methods (in)consistency?', *Experimental Economics* **24**(2), 593–616.
URL: *https://doi.org/10.1007/s10683-020-09674-8*

Hommes, C. (2013), Laboratory experiments, *in* 'Behavioral Rationality and Heterogeneous Expectations in Complex Economic Systems', Cambridge University Press, pp. 211–236.
URL: *https://doi.org/10.1017/cbo9781139094276.009*

Horowitz, J. K. (2006), 'The becker-DeGroot-marschak mechanism is not necessarily incentive compatible, even for non-random goods', *Economics Letters* **93**(1), 6–11.
URL: *https://doi.org/10.1016/j.econlet.2006.03.033*

Huang, C.-C. (2006), Cheating by men in the gale-shapley stable matching algorithm, *in* Y. Azar and T. Erlebach, eds, 'Algorithms – ESA 2006', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 418–431.

Hurley, T. M., Peterson, N. R. and Shogren, J. F. (2007), Belief elicitation: An experimental comparison of scoring rule and prediction methods.

Hurley, T. M. and Shogren, J. F. (2005), 'An experimental comparison of induced and elicited beliefs', *Journal of Risk and Uncertainty* **30**(2), 169–188.
URL: *https://doi.org/10.1007/s11166-005-6565-5*

Hvide, H., Lee, J. and Odean, T. (2019), 'Easy money, cheap talk, or spuds: Inducing risk aversion in economics experiments', *SSRN Electronic Journal* .
URL: *https://doi.org/10.2139/ssrn.3433380*

Imai, K. (2011), 'Multivariate regression analysis for the item count technique', *Journal of the American Statistical Association* **106**(494), 407–416.
URL: *https://doi.org/10.1198/jasa.2011.ap10415*

Imamura, K. (2021), 'Meritocracy versus diversity'.

URL: *https://drive.google.com/file/d/1zF2cPq7BmDgrDgkV6PoPHqGIwBZ-Uhrf/view*

Jin, J., He, R., Gong, H., Xu, X. and He, C. (2017), 'Farmers' risk preferences in rural china: Measurements and determinants', *International Journal of Environmental Research and Public Health* **14**(7), 713.
URL: *https://doi.org/10.3390/ijerph14070713*

Johnson, C., Baillon, A., Bleichrodt, H., Li, Z., van Dolder, D. and Wakker, P. P. (2021), 'Prince: An improved method for measuring incentivized preferences', *Journal of Risk and Uncertainty* **62**(1), 1–28.
URL: *https://doi.org/10.1007/s11166-021-09346-9*

Johnston, R. J., Boyle, K. J., Adamowicz, W. V., Bennett, J., Brouwer, R., Cameron, T. A., Hanemann, W. M., Hanley, N., Ryan, M., Scarpa, R., Tourangeau, R. and Vossler, C. A. (2017), 'Contemporary guidance for stated preference studies', *Journal of the Association of Environmental and Resource Economists* **4**(2), 319–405.
URL: *https://doi.org/10.1086/691697*

Kahneman, D., Knetsch, J. L. and Thaler, R. H. (1986), 'Fairness and the assumptions of economics', *The Journal of Business* **59**(S4), S285.
URL: *https://doi.org/10.1086/296367*

Kahneman, D. and Tversky, A. (1979), 'Prospect theory: An analysis of decision under risk', *Econometrica* **47**(2), 263.
URL: *https://doi.org/10.2307/1914185*

Karni, E. and Safra, Z. (1987), '"preference reversal" and the observability of preferences by experimental methods', *Econometrica* **55**(3), 675.
URL: *https://doi.org/10.2307/1913606*

Kelso, A. S. and Crawford, V. P. (1982), 'Job matching, coalition formation, and gross substitutes', **50**(6), 1483.
URL: *https://doi.org/10.2307/1913392*

Kessler, J. B. and Vesterlund, L. (2015), The external validity of laboratory experiments: The misleading emphasis on quantitative effects, *in* 'Handbook of Experimental Economic Methodology', Oxford University Press, pp. 391–406.
URL: *https://doi.org/10.1093/acprof:oso/9780195328325.003.0020*

Kesten, O. (2010), 'School choice with consent*', *The Quarterly Journal of Eco-*

nomics **125**(3), 1297–1348.
URL: *http://dx.doi.org/10.1162/qjec.2010.125.3.1297*

Kesten, O. and Ünver, M. U. (2015), 'A theory of school choice lotteries', *Theoretical Economics* **10**(2).
URL: *https://ideas.repec.org/a/the/publsh/1558.html*

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S. and Saez, E. (2011), 'Unwilling or unable to cheat? evidence from a tax audit experiment in denmark', *Econometrica* **79**(3), 651–692.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA9113*

Knight, F. H. (1921), *Risk, Uncertainty and Profit*, Houghton Mifflin Company.
URL: *https://fraser.stlouisfed.org/files/docs/publications/books/risk/riskuncertaintyprofit.pdf*

Koch, A. K. and Normann, H.-T. (2008), 'Giving in dictator games: Regard for others or regard by others?', *Southern Economic Journal* **75**(1), 223–231.
URL: *http://www.jstor.org/stable/20112036*

Kojima, F. and Ünver, M. U. (2014), 'The "boston" school-choice mechanism: an axiomatic approach', *Economic Theory* **55**(3), 515–544.
URL: *http://www.jstor.org/stable/43562942*

Konow, J. (2000), 'Fair shares: Accountability and cognitive dissonance in allocation decisions', *American Economic Review* **90**(4), 1072–1091.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.90.4.1072*

Konow, J. (2010), 'Mixed feelings: Theories of and evidence on giving', *Journal of Public Economics* **94**(3), 279–297.
URL: *https://www.sciencedirect.com/science/article/pii/S0047272709001480*

Korenok, O., Millner, E. L. and Razzolini, L. (2018), 'Taking aversion', *Journal of Economic Behavior & Organization* **150**, 397–403.
URL: *https://doi.org/10.1016/j.jebo.2018.01.021*

Larney, A., Rotella, A. and Barclay, P. (2019), 'Stake size effects in ultimatum game and dictator game offers: A meta-analysis', *Organizational Behavior and Human Decision Processes* **151**, 61–72.
URL: *https://doi.org/10.1016/j.obhdp.2019.01.002*

Lee, K. S. (2005), Rationality, minds, and machines in the laboratory: A thematic history of Vernon Smith's experimental economics, PhD thesis, University of Notre Dame, Ann Arbor.
URL: *https://www.proquest.com/docview/305422455/abstract/506D573BDA5D49E4PQ/*

Lee, V. E. and Bryk, A. S. (1989), 'A multilevel model of the social distribution of high school achievement', *Sociology of Education* **62**(3), 172–192.
URL: *http://www.jstor.org/stable/2112866*

Lee, V. E., Bryk, A. S. and Smith, J. B. (1993), 'The organization of effective secondary schools', *Review of Research in Education* **19**, 171–267.
URL: *http://www.jstor.org/stable/1167343*

Lenfant, J.-S. (2012), 'Indifference curves and the ordinalist revolution', *History of Political Economy* **44**(1), 113–155.
URL: *https://doi.org/10.1215/00182702-1504077*

Leshno, J. (2021), 'Stable matching with peer-dependent preferences in large markets: Existence and cutoff characterization', *SSRN Electronic Journal* .

Levitt, S. D. and List, J. A. (2009), 'Field experiments in economics: The past, the present, and the future', *European Economic Review* **53**(1), 1–18.
URL: *https://doi.org/10.1016/j.euroecorev.2008.12.001*

Lewis, L., Snow, K., Farris, E., Smerdon, B., Cronen, S. and Kaplan, J. (2000), Condition of america's public school facilities: 1999, Technical report, U.S. Department of Education, National Center for Education Statistics.
URL: *https://nces.ed.gov/pubs2000/2000032.pdf*

Lezzi, E., Fleming, P. and Zizzo, D. J. (2015), Does it matter which effort task you use? A comparison of four effort tasks when agents compete for a prize, Working Paper series, University of East Anglia, Centre for Behavioural and Experimental Social Science (CBESS) 15-05, School of Economics, University of East Anglia, Norwich, UK.
URL: *https://ideas.repec.org/p/uea/wcbess/15-05.html*

Li, Y. (2016), Incentive Compatible Payoff Mechanisms for General Risk Theories, PhD thesis, Georgia State University.
URL: *https://scholarworks.gsu.edu/econ_diss/120/*

Liebe, U. and Tutic, A. (2010), 'Status groups and altruistic behaviour in dictator games', *Rationality and Society* **22**(3), 353–380.
URL: *https://doi.org/10.1177/1043463110366232*

List, J. A. (2007), 'On the interpretation of giving in dictator games', *Journal of Political Economy* **115**(3), 482–493.
URL: *https://doi.org/10.1086/519249*

List, J. A. and Lucking-Reiley, D. (2002), 'The effects of seed money and

refunds on charitable giving: Experimental evidence from a university capital campaign', *Journal of Political Economy* **110**(1), 215–233.
URL: *https://doi.org/10.1086/324392*

Loomes, G. (1999), 'Some lessons from past experiments and some challenges for the future', *The Economic Journal* **109**(453), F35–F45.
URL: *http://www.jstor.org/stable/2565584*

Loomes, G. and Segal, U. (1994), 'Observing different orders of risk aversion', *Journal of Risk and Uncertainty* **9**(3), 239–256.
URL: *http://www.jstor.org/stable/41760748*

Machina, M. J. (1989), 'Dynamic consistency and non-expected utility models of choice under uncertainty', *Journal of Economic Literature* **27**(4), 1622–1668.
URL: *http://www.jstor.org/stable/2727025*

March, C., Ziegelmeyer, A., Greiner, B. and Cyranek, R. (2016), Pay Few Subjects but Pay Them Well: Cost-Effectiveness of Random Incentive Systems, Technical report.

Marlowe, F. (2004), Dictators and ultimatums in an egalitarian society of hunter-gatherers: The hadza of tanzania, *in* 'Foundations of Human Sociality', Oxford University Press, pp. 168–193.
URL: *https://doi.org/10.1093/0199262055.003.0006*

Massar, S. A. A., Csathó, r. and Van der Linden, D. (2018), 'Quantifying the motivational effects of cognitive fatigue through effort-based decision making.', *Frontiers in psychology* **9**, 843.

McDermott, R. (2011), 'New directions for experimental work in international relations', *International Studies Quarterly* **55**(2), 503–520.
URL: *http://www.jstor.org/stable/23019700*

McVitie, D. G. and Wilson, L. B. (1971), 'The stable marriage problem', *Communications of the ACM* **14**(7), 486–490.
URL: *https://doi.org/10.1145/362619.362631*

Meijers, M. H., Verlegh, P. W., Noordewier, M. K. and Smit, E. G. (2015), 'The dark side of donating: how donating may license environmentally unfriendly behavior', *Social Influence* **10**(4), 250–263.
URL: *https://doi.org/10.1080/15534510.2015.1092468*

Merlo, A. and Schotter, A. (2003), 'Learning by not doing: an experimen-

tal investigation of observational learning', *Games and Economic Behavior* **42**(1), 116–136.
URL: *https://doi.org/10.1016/s0899-8256(02)00537-7*

Miao, B. and Zhong, S. (2016), 'Probabilistic social preference: how machina's mom randomizes her choice', *Economic Theory* **65**(1), 1–24.
URL: *https://doi.org/10.1007/s00199-016-1015-y*

Miguel, E. and Kremer, M. (2004), 'Worms: Identifying impacts on education and health in the presence of treatment externalities', *Econometrica* **72**(1), 159–217.
URL: *https://doi.org/10.1111/j.1468-0262.2004.00481.x*

Miller, J. D. (1984), A new survey technique for studying deviant behavior, Thesis, George Washington University.

Mosteller, F. and Nogee, P. (1951), 'An experimental measurement of utility', *Journal of Political Economy* **59**(5), 371–404.
URL: *http://www.jstor.org/stable/1825254*

Mujcic, R. and Frijters, P. (2011), Altruism in society: Evidence from a natural experiment involving commuters, IZA Discussion Papers 5648, Institute of Labor Economics (IZA).
URL: *https://EconPapers.repec.org/RePEc:iza:izadps:dp5648*

Mullette-Gillman, O. A., Leong, R. L. F. and Kurnianingsih, Y. A. (2015), 'Cognitive fatigue destabilizes economic decision making preferences and strategies', *PLOS ONE* **10**(7), 1–19.
URL: *https://doi.org/10.1371/journal.pone.0132022*

Neill, H. R., Cummings, R. G., Ganderton, P. T., Harrison, G. W. and McGuckin, T. (1994), 'Hypothetical surveys and real economic commitments', *Land Economics* **70**(2), 145.
URL: *https://doi.org/10.2307/3146318*

Niederle, M., Roth, A. E. and Sönmez, T. (2008), *Matching*, 2 edn, Palgrave Macmillan, p. 434–444.

Niederle, M. and Vesterlund, L. (2007), 'Do women shy away from competition? do men compete too much?', *The Quarterly Journal of Economics* **122**(3), 1067–1101.
URL: *https://doi.org/10.1162/qjec.122.3.1067*

Noussair, C. N. and van Soest, D. P. (2014), 'Economic experiments and en-

vironmental policy', *Annual Review of Resource Economics* **6**(1), 319–337.
URL: *https://doi.org/10.1146/annurev-resource-091912-151833*

Null, C. (2011), 'Warm glow, information, and inefficient charitable giving', *Journal of Public Economics* **95**(5-6), 455–465.
URL: *https://doi.org/10.1016/j.jpubeco.2010.06.018*

O'Donoghue, T. and Somerville, J. (2018), 'Modeling risk aversion in economics', *Journal of Economic Perspectives* **32**(2), 91–114.
URL: *https://doi.org/10.1257/jep.32.2.91*

Ottoni-Wilhelm, M., Vesterlund, L. and Xie, H. (2017), 'Why do people give? testing pure and impure altruism', *American Economic Review* **107**(11), 3617–3633.
URL: *https://doi.org/10.1257/aer.20141222*

Pablo, B.-G., Bucheli, M., Espinosa, M. P. and Teresa, G.-M. (2013), 'Moral cleansing and moral licenses: Experimental evidence', *Economics and Philosophy* **29**(2), 199–212.
URL: *https://EconPapers.repec.org/RePEc:cup:ecnphi:v:29:y:2013:i:02:p:199-212_00*

Padoa-Schioppa, C. (2011), 'Neurobiology of economic choice: A good-based model', *Annual Review of Neuroscience* **34**(1), 333–359. PMID: 21456961.
URL: *https://doi.org/10.1146/annurev-neuro-061010-113648*

Palfrey, T. R. (2009), 'Laboratory experiments in political economy', *Annual Review of Political Science* **12**(1), 379–388.
URL: *https://doi.org/10.1146/annurev.polisci.12.091007.122139*

Pathak, P. A. (2017), *What Really Matters in Designing School Choice Mechanisms*, Vol. 1 of *Econometric Society Monographs*, Cambridge University Press, p. 176–214.

Peters, J. and Büchel, C. (2009), 'Overlapping and distinct neural systems code for subjective value during intertemporal and risky decision making', *Journal of Neuroscience* **29**(50), 15727–15734.
URL: *https://www.jneurosci.org/content/29/50/15727*

Phan, W., Tierney, R. and Zhou, Y. (2021), Crowding in school choice, Technical report.

Ploner, M. and Regner, T. (2013), Self-Image and Moral Balancing - An Ex-

perimental Analysis, Jena Economic Research Papers 2013-002, Friedrich-Schiller-University Jena.
URL: *https://ideas.repec.org/p/jrp/jrpwrp/2013-002.html*

Ponti, G. and Rodriguez-Lara, I. (2015), 'Social preferences and cognitive reflection: evidence from a dictator game experiment', *Frontiers in Behavioral Neuroscience* **9**.
URL: *https://doi.org/10.3389/fnbeh.2015.00146*

Predmore, C., Topyan, K. and Apadula, L. T. (2021), 'Impact of process misconception in becker-DeGroot-marschak single response value elicitation procedures: An experimental investigation in consumer behavior using the IKEA effect', *Economies* **9**(4), 173.
URL: *https://doi.org/10.3390/economies9040173*

Pycia, M. (2012), 'Stability and preference alignment in matching and coalition formation', *Econometrica* **80**(1), 323–362.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA7143*

Pycia, M. and Ünver, M. U. (2011), 'Trading cycles for school choice'.

Pycia, M. and Yenmez, M. B. (2019), Matching with Externalities, CEPR Discussion Papers 13994, C.E.P.R. Discussion Papers.
URL: *https://ideas.repec.org/p/cpr/ceprdp/13994.html*

Quiggin, J. (1982), 'A theory of anticipated utility', *Journal of Economic Behavior & Organization* **3**(4), 323–343.
URL: *https://doi.org/10.1016/0167-2681(82)90008-7*

Rabin, M. (1993), 'Incorporating fairness into game theory and economics', *The American Economic Review* **83**(5), 1281–1302.
URL: *http://www.jstor.org/stable/2117561*

Rabin, M. (2000), 'Risk aversion and expected-utility theory: A calibration theorem', *Econometrica* **68**(5), 1281–1292.
URL: *http://www.jstor.org/stable/2999450*

Rand, D. G. (2016), 'Cooperation, fast and slow', **27**(9), 1192–1206.
URL: *https://doi.org/10.1177/0956797616654455*

Ready, D. D., Lee, V. E. and Welner, K. G. (2004), 'Educational equity and school structure: School size, overcrowding, and schools-within-schools', *Teachers College Record* **106**(10), 1989–2014.
URL: *https://nepc.colorado.edu/sites/default/files/1882.pdf*

Reteig, L. C., van den Brink, R. L., Prinssen, S., Cohen, M. X. and Slagter, H. A. (2019), 'Sustaining attention for a prolonged period of time increases temporal variability in cortical responses', *Cortex* **117**, 16 – 32.
URL: *http://www.sciencedirect.com/science/article/pii/S0010945219300681*

Roth, A. (1986), 'On the allocation of residents to rural hospitals: A general property of two-sided matching markets', *Econometrica* **54**(2), 425–27.
URL: *https://EconPapers.repec.org/RePEc:ecm:emetrp:v:54:y:1986:i:2:p:425-27*

Roth, A. E. (1982), 'The Economics of Matching: Stability and Incentives', *Mathematics of Operations Research* **7**(4), 617–628.
URL: *https://ideas.repec.org/a/inm/ormoor/v7y1982i4p617-628.html*

Roth, A. E. (1991), 'Game theory as a part of empirical economics', *The Economic Journal* **101**(404), 107.
URL: *https://doi.org/10.2307/2233845*

Roth, A. E. (1993), 'The early history of experimental economics', *Journal of the History of Economic Thought* **15**(2), 184–209.
URL: *https://doi.org/10.1017/s1053837200000936*

Roth, A. E. (2016), 5. experiments in market design, *in* J. H. Kagel and A. E. Roth, eds, 'The Handbook of Experimental Economics, Volume Two', Princeton University Press.
URL: *https://doi.org/10.1515/9781400883172-006*

Roth, A. E. and Murnighan, J. (1978), 'Equilibrium behavior and repeated play of the prisoner's dilemma', *Journal of Mathematical Psychology* **17**(2), 189 – 198.
URL: *http://www.sciencedirect.com/science/article/pii/0022249678900305*

Roth, A. E. and Wilson, R. B. (2019), 'How market design emerged from game theory: A mutual interview', *Journal of Economic Perspectives* **33**(3), 118–143.
URL: *https://doi.org/10.1257/jep.33.3.118*

Rousseas, S. W. and Hart, A. G. (1951), 'Experimental verification of a composite indifference map', *Journal of Political Economy* **59**(4), 288–318.
URL: *https://doi.org/10.1086/257092*

Sadiraj, V. (2013), 'Probabilistic risk attitudes and local risk aversion: a paradox', *Theory and Decision* **77**(4), 443–454.
URL: *https://doi.org/10.1007/s11238-013-9410-3*

Safra, Z. and Segal, U. (2008), ‘Calibration results for non-expected utility theories’, *Econometrica* **76**(5), 1143–1166.
URL: *http://www.jstor.org/stable/40056496*

Safra, Z., Segal, U. and Spivak, A. (1990), ‘The becker-degroot-marschak mechanism and nonexpected utility: A testable approach’, *Journal of Risk and Uncertainty* **3**(2), 177–190.
URL: *http://www.jstor.org/stable/41760593*

Saglam, I. and Mumcu, A. (2007), ‘The core of a housing market with externalities’, *Economics Bulletin* **3**(57), 1–5.
URL: *https://ideas.repec.org/a/ebl/ecbull/eb-07c70026.html*

Samuelson, L. (2005), ‘Economic theory and experimental economics’, *Journal of Economic Literature* **43**(1), 65–107.
URL: *http://www.jstor.org/stable/4129307*

Samuelson, P. A. (1950), ‘The problem of integrability in utility theory’, *Economica* **17**(68), 355.
URL: *https://doi.org/10.2307/2549499*

Sasaki, H. and Toda, M. (1996), ‘Two-sided matching problems with externalities’, *Journal of Economic Theory* **70**(1), 93–108.
URL: *https://EconPapers.repec.org/RePEc:eee:jetheo:v:70:y:1996:i:1:p:93-108*

Schotter, A. and Trevino, I. (2014), ‘Belief elicitation in the laboratory’, *Annual Review of Economics* **6**(1), 103–128.
URL: *https://doi.org/10.1146/annurev-economics-080213-040927*

Schouppe, N., Ridderinkhof, K. R., Verguts, T. and Notebaert, W. (2014), ‘Context-specific control and context selection in conflict tasks’, *Acta Psychologica* **146**, 63–66.
URL: *https://doi.org/10.1016/j.actpsy.2013.11.010*

Schram, A. (2005), ‘Artificiality: The tension between internal and external validity in economic experiments’, *Journal of Economic Methodology* **12**(2), 225–237.
URL: *https://doi.org/10.1080/13501780500086081*

Schurter, K. and Wilson, B. J. (2009), ‘Justice and fairness in the dictator game’, *Southern Economic Journal* **76**(1), 130–145.
URL: *http://www.jstor.org/stable/27751456*

Segal, U. (1988), ‘Does the preference reversal phenomenon necessar-

ily contradict the independence axiom?', *The American Economic Review* **78**(1), 233–236.
URL: *http://www.jstor.org/stable/1814711*

Segal, U. (1990), 'Two-stage lotteries without the reduction axiom', *Econometrica* **58**(2), 349.
URL: *https://doi.org/10.2307/2938207*

Segal, U. (1992), The independence axiom versus the reduction axiom: Must we have both?, *in* 'Utility Theories: Measurements and Applications', Springer Netherlands, pp. 165–183.
URL: *https://doi.org/10.1007/978-94-011-2952-7_7*

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D. and Botvinick, M. M. (2017), 'Toward a rational and mechanistic account of mental effort', *Annual Review of Neuroscience* **40**(1), 99–124. PMID: 28375769.
URL: *https://doi.org/10.1146/annurev-neuro-072116-031526*

Sherstyuk, K., Tarui, N. and Saijo, T. (2013), 'Payment schemes in infinite-horizon experimental games', *Experimental Economics* **16**(1), 125–153.
URL: *https://doi.org/10.1007/s10683-012-9323-y*

Sims, C. A. (2003), 'Implications of rational inattention', *Journal of Monetary Economics* **50**(3), 665–690.
URL: *https://doi.org/10.1016/s0304-3932(03)00029-1*

Smith, V. L. and Walker, J. M. (1993), 'Monetary rewards and decision cost in experimental economics', *Economic Inquiry* **31**(2), 245–261.
URL: *https://doi.org/10.1111/j.1465-7295.1993.tb00881.x*

Sönmez, T. (2013), 'Bidding for army career specialties: Improving the rotc branching mechanism', *Journal of Political Economy* **121**(1), 186–219.
URL: *https://doi.org/10.1086/669915*

Sönmez, T. and Switzer, T. (2013), 'Matching with (branch-of-choice) contracts at the united states military academy', **81**(2), 451–488.
URL: *https://doi.org/10.3982/ecta10570*

Sönmez, T. and Ünver, U. (2009), Matching, allocation, and exchange of discrete resources, Boston College Working Papers in Economics 717, Boston College Department of Economics.
URL: *https://EconPapers.repec.org/RePEc:boc:bocoec:717*

Sönmez, T. and Ünver, U. (2010), 'Course bidding at business schools*', *International Economic Review* **51**(1), 99–123.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-2354.2009.00572.x*

Starmer, C. and Sugden, R. (1991), 'Does the random-lottery incentive system elicit true preferences? an experimental investigation', *The American Economic Review* **81**(4), 971–978.
URL: *http://www.jstor.org/stable/2006657*

Stewart, N., Canic, E. and Mullett, T. L. (2019), 'On the futility of estimating utility functions: Why the parameters we measure are wrong, and why they do not generalize'.
URL: *https://doi.org/10.31234/osf.io/qt69m*

Thurstone, L. L. (1931), 'The indifference function', *The Journal of Social Psychology* **2**(2), 139–167.
URL: *https://doi.org/10.1080/00224545.1931.9918964*

Torres, C., Hanley, N. and Riera, A. (2011), 'How wrong can you be? implications of incorrect utility function specification for welfare measurement in choice experiments', *Journal of Environmental Economics and Management* **62**(1), 111–121.
URL: *https://doi.org/10.1016/j.jeem.2010.11.007*

Troyan, P. (2012), 'Comparing school choice mechanisms by interim and ex-ante welfare', *Games and Economic Behavior* **75**(2), 936 – 947.
URL: *http://www.sciencedirect.com/science/article/pii/S0899825612000103*

Tversky, A. and Kahneman, D. (1992), 'Advances in prospect theory: Cumulative representation of uncertainty', *Journal of Risk and Uncertainty* **5**(4), 297–323.
URL: *http://www.jstor.org/stable/41755005*

Umemoto, A., Inzlicht, M. and Holroyd, C. B. (2019), 'Electrophysiological indices of anterior cingulate cortex function reveal changing levels of cognitive effort and reward valuation that sustain task performance', *Neuropsychologia* **123**, 67 – 76. Cognitive Effort.
URL: *http://www.sciencedirect.com/science/article/pii/S0028393218302823*

Umer, H., Kurosaki, T. and Iwasaki, I. (2022), 'Unearned endowment and charity recipient lead to higher donations: A meta-analysis of the dictator

game lab experiments', *Journal of Behavioral and Experimental Economics* **97**, 101827.
URL: *https://doi.org/10.1016/j.socec.2022.101827*

Vass, C., Rigby, D. and Payne, K. (2017), 'The role of qualitative research methods in discrete choice experiments', *Medical Decision Making* **37**(3), 298–313.
URL: *https://doi.org/10.1177/0272989x16683934*

Vieider, F. M., Lefebvre, M., Bouchouicha, R., Chmura, T., Hakimov, R., Krawczyk, M. and Martinsson, P. (2014), 'Common components of risk and uncertainty attitudes across contexts and domains: Evidence from 30 countries', *Journal of the European Economic Association* **13**(3), 421–452.
URL: *https://doi.org/10.1111/jeea.12102*

Villeval, M.-C., Gupta, N. D. and Poulsen, A. (2005), Male and Female Competitive Behavior - Experimental Evidence, Working Papers 0512, Groupe d'Analyse et de Théorie Economique Lyon St-Étienne (GATE Lyon St-Étienne), Université de Lyon.
URL: *https://ideas.repec.org/p/gat/wpaper/0512.html*

Walasek, L. and Stewart, N. (2015), 'How to make loss aversion disappear and reverse: Tests of the decision by sampling origin of loss aversion.', *Journal of Experimental Psychology: General* **144**(1), 7–11.
URL: *https://doi.org/10.1037/xge0000039*

Walasek, L. and Stewart, N. (2021), 'You cannot accurately estimate an individual's loss aversion using an accept-reject task.', *Decision* **8**(1), 2–15.
URL: *https://doi.org/10.1037/dec0000141*

Wallis, W. A. and Friedman, M. (1942), *The Empirical Derivation of Indifference Functions*, University of Chicago Press, pp. 175–89.

Wang, S. W. (2011), 'Incentive effects: The case of belief elicitation from individuals in groups', *Economics Letters* **111**(1), 30–33.
URL: *https://doi.org/10.1016/j.econlet.2010.11.045*

Wang, X. and Navarro-Martinez, D. (2019), Bridging the Gap Between the Lab and the Field: Dictator Games and Donations, Working paper, Pompeu Fabra University; Barcelona Graduate School of Economics.
URL: *https://www.upf.edu/documents/2963149/229409653/JMP+Xinghua+Wang.pdf/ff52c394-e2f8-c4d8-994d-2e6374471453*

WCPSS (2017), *2016-2017 Facilities Utilization Report*, WCPSS.
URL: *https://www.wcpss.net/cms/lib/NC01911451/Centricity/Domain/100/2016-17-facilities-utilization-report with_ADM.pdf*

White, C. J., Kelly, J. M., Shariff, A. F. and Norenzayan, A. (2019), 'Supernatural norm enforcement: Thinking about karma and god reduces selfishness among believers', *Journal of Experimental Social Psychology* **84**, 103797.
URL: *https://doi.org/10.1016/j.jesp.2019.03.008*

Wilcox, N. T. (2015), Unusual Estimates of Probability Weighting Functions, Technical report.

Wu, S., Sun, S., Camilleri, J. A., Eickhoff, S. B. and Yu, R. (2021), 'Better the devil you know than the devil you don't: Neural processing of risk and ambiguity', *NeuroImage* **236**, 118109.
URL: *https://doi.org/10.1016/j.neuroimage.2021.118109*

Wylie, G. R., Genova, H. M., DeLuca, J. and Dobryakova, E. (2017), 'The relationship between outcome prediction and cognitive fatigue: A convergence of paradigms', *Cognitive, Affective, & Behavioral Neuroscience* **17**(4), 838–849.
URL: *https://doi.org/10.3758/s13415-017-0515-y*

Yaari, M. E. (1987), 'The dual theory of choice under risk', *Econometrica* **55**(1), 95.
URL: *https://doi.org/10.2307/1911158*

Yang, C.-L. and Yao, L. (2017), 'Testing ambiguity theories with a mean-preserving design', *Quantitative Economics* **8**(1), 219–238.
URL: *https://doi.org/10.3982/qe460*

Zhang, Y. J. (2013), 'Can experimental economics explain competitive behavior outside the lab?', *SSRN Electronic Journal* .
URL: *https://doi.org/10.2139/ssrn.2292929*

Zhou, W. and Hey, J. (2018), 'Context matters', *Experimental Economics* **21**(4), 723–756.
URL: *https://doi.org/10.1007/s10683-017-9546-z*

Zizzo, D. J. (2009), 'Experimenter demand effects in economic experiments', **13**(1), 75–98.
URL: *https://doi.org/10.1007/s10683-009-9230-z*

5

Appendices for Chapter 1: Compensation Without Distortion

A Properties

Here, I define the minimum necessary assumptions on $\succeq$ such that the random stopping procedure is incentive compatible. I first introduce some additional notation that draws heavily from Azrieli et al. (2018), before demonstrating key properties.

A.1 Notation

Continue to use $d_i(A_i, B_i)$ to denote a decision question that presents two options A_i and B_i in round i. A payment procedure can be defined as a mechanism ϕ that maps responses from $d_1, \ldots, d_n$ onto a payoff object, denoted $\mathcal{L}$. Subjects are asked to make choices over a series of questions d_i, however have preferences over $\mathcal{L}$. Failures of incentive compatibility, then, can intuitively be thought of as circumstances when a subject prefers some A_i to B_i in isolation, however because of the definition of ϕ, prefers $\mathcal{L}(B_i)$ over $\mathcal{L}(A_i)$.

Generally, it is easiest define $\mathcal{L}$ using lottery notation. If a subject were to receive some outcome x if event v occurred, and some outcome y if event ζ occurred, the payoff object could be written as: $\mathcal{L}(x, v;\ y, \zeta)$.

To provide a bit more clarity, consider a two question experiment. If the subject is given their choices in both rounds (corresponding to a pay all mechanism), then the payoff object $\mathcal{L}$ would be defined as $\mathcal{L}(\{c(d_1) \cap$

$c(d_2)\}, 1)$. If, instead, the participant was given either decision with probability 1/2, the payoff object for this two round experiment would be:

$$\mathcal{L}(c(d_1),\ 1/2;\ c(d_2),\ 1/2)$$

A.2 Incentive Compatibility

First, consider the minimum necessary assumptions on $\succeq$ such that the random stopping procedure is incentive compatible.

Consider an individual facing a decision in round $i < n$. Given the individual is answering $d_i(A_i, B_i)$, it must be the case that no previous decision has been paid. Denote some future decision being paid by ε, and note that the incentives for reporting $c(d_i)$ as $A_i \succ B_i$ is:

$$(A_i, p_i;\ \varepsilon, 1 - p_i) \succeq (B_i, p_i;\ \varepsilon, 1 - p_i) \tag{4}$$

If preferences $\succeq$ satisfy assumption 1, then

$$(A_i, p_i;\ \varepsilon, 1 - p_i) \succeq (B_i, p_i;\ \varepsilon, 1 - p_i) \iff A_i \succeq B_i \tag{5}$$

Alternatively, note that ε is structurally equivalent to the event that neither A_i nor B_i are compensated. Equation 5 can thus be rewritten to:

$$(A_i, p_i;\ \neg c(d_i), 1 - p_i) \succeq (B_i, p_i;\ \neg c(d_i), 1 - p_i) \iff A_i \succeq B_i \tag{6}$$

A.3 Portfolio Effects

Next, I turn to demonstrating the lack of portfolio effects in an RSP experiment. Using $\mathcal{L}$ notation, what is necessary for portfolio effects to exist is for multiple decisions to coexist in the same experimental outcome. In other words, if there exist two decisions d_i and d_j such that there exists an experiment outcome ω where $Pr(\mathcal{L} = d_i \cap d_j) \neq 0$. Note, this condition is excluded due to the existence of the termination lottery.

A.4 Wealth or Income Effects

The termination lottery excludes endogenously-generated wealth effects for reasons similar to those mentioned in section A.3 above. Formally, suppose an experiment is sensitive to endogenous wealth effects. By definition, this requires there to be some round after which a subject's contemporaneous level of wealth has changed. Define a subject's experimentally-induced wealth (the individual's pre-experiment level of wealth plus income generated by experimental history ω) as $\mathcal{I}(\omega)$. Without lack of generality, consider some decision round j, where $1 < j < n$, such that:

$$\mathbb{E}[\mathcal{I}_{\omega_{1,\ldots,j}}] \neq \mathbb{E}[\mathcal{I}_{\omega_{j,\ldots,n}}] \tag{7}$$

Note that equation 7 requires the existence of some j such that, after decision d_j, a subject's expectations regarding their income have changed. In a random stopping procedure experiment, the termination lottery removes this possibility. In real terms, there does not exist any ω where a d_j is compensated and any decision $d_{i>j}$ is observed. In addition, assumption 1 removes the possibility of a change in subject expectations regarding their experimental payoff.

B More Information on Recruitment Sensitivity

To better contextualize the impact of different parameters, table A.1 gives various experimental parameter values and the associated initial recruitment pool needed.

Table A.1: Necessary recruitment pool (N) given specific experimental variables and desired sample size.

Threshold Round n	CCP p value	Implied ρ_n	Desired $\mathcal{S}$	Confidence Level	Necessary N
10	.025	.796	100	.975	137
10	.05	.63	100	.975	179
10	.05	.63	100	.99	183
10	.10	.39	100	.975	301
20	.05	.38	100	.975	310
50	.015	.477	100	.975	242
50	.025	.289	100	.975	407

B.1 k-answers Per Question

In this context, the necessary number of recruited subjects additionally depends on how many distinct lab sessions are held. Every "new session" presents an opportunity to reshuffle the question order, pushing the least answered questions to the front of the queue. The more sessions, the smaller N can be for any level of certainty. This is most straightforwardly observed by considering the theoretically optimal "N session" procedure: the first subject begins with question 1 and continues until stopped by the termination lottery—denoted question i. The second subject then begins with question $i+1$, and continues in the standard order (after question n, moving to question 1, if not stopped). The "n groups of N/n" method is the most efficient construction without any reshuffling, and is therefore an upper-bound estimate for necessary recruitment.

B.2 Cost Estimates and Additional Details

To compare the RSP and other methods, I take several previously published experiments and estimate each one's expected cost under a pay one randomly, RSP, and pay all design. Experiments were selected to represent a relatively broad cross-section while remaining similar enough in design and composition to be generally comparable. Specifically, each experiment could be reduced to a series of n binary-choice decision questions, where the expected value for each choice was calculable. For every experiment, expected experimental costs are presented as a range: $\mathbb{E}_{min}$ gives the expected cost if every subject always picked the lower expected value option in every round, while $\mathbb{E}_{max}$ assumes all individuals always choose the higher expected value choice option. The results are given in table A.2.

Table A.2: Comparison of expected costs across different payment procedures (pay one randomly, random stopping procedure, pay all, and one task) for selection of applicable historical experiments.

	POR		RSP		PA		OT[a]	
Experiment	$\mathbb{E}_{\mathbf{min}}$	$\mathbb{E}_{\mathbf{max}}$	$\mathbb{E}_{\mathbf{min}}$	$\mathbb{E}_{\mathbf{max}}$	$\mathbb{E}_{\mathbf{min}}$	$\mathbb{E}_{\mathbf{max}}$	$\mathbb{E}_{\mathbf{min}}$	$\mathbb{E}_{\mathbf{max}}$
Harrison and Swarthout (2014)[b]	6,349	7,826	16,973	20,921	145,238	189,547	190,478	234,787
Loomes, Starmer and Sugden (1991) (Sub-sample 1)[c]	376	418	998	1,111	7,527	8,379	*7,527*	*8,379*
Cox, Sadiraj, and Schmidt (2015)[d]	294	336	721	824	1,470	1,680	*1,470*	*1,680*
Hey and Lee (2005)[e]	4,773	5,683	12,773	15,208	143,200	170,497	*143,200*	*170,497*

[a]For experiments with no mention of whether subjects were given a "show up" payment, the PA and OT costs are the same. In these cases, the OT cost columns are italicized to highlight the exclusion of any show up payment in cost calculations.
[b]I use the N=208 completions associated with Harrison and Swarthout (2014)'s 1-in-30 sample for cost estimates.
[c]Using the sub-sample 2, who were asked slightly different questions, lowers all methods' minimum expected cost to run by 3.7% and raises all methods' maximum expected cost by 2.5%
[d]Cox Sadiraj and Schmidt (2015) use different N values for different procedures. For these cost estimates, I assume 40 necessary completions which corresponds with their most frequent N.
[e]Hey and Lee (2005) use two different procedures with different groups—I only consider the 179-subject pairwise choice group.

This exercise demonstrates expected cost for an RSP experiment rests between the POR lower bound, and the PA/OT upper bound. Moreover, while the increase in costs is relatively uniform moving from a POR to RSP experiment (the RSP is between 2.45 and 2.68 times as expensive as its POR equivalent across all comparisons), the savings the RSP offers compared to a PA or OT procedure varies more widely. For Cox et al. (2015), the PA design is only twice as expensive as an RSP equivalent, while Hey and Lee (2005)'s experiment would be more than eleven times as expensive to run as a PA over an RSP.

The reasons for these differences highlight the specific dependencies of each procedure on prominent design characteristics. Relatively straightforwardly, POR costs are a function of the number of subjects (N) and the expected decision-dependent payment. RSP experimental costs depend on the same two factors, but also generally require a larger pool of subjects to ensure a sufficient of completions.[119] PA experimental costs depend on the expected sum of all round costs, and are thus most sensitive to both average round payouts and the number of decision rounds presented. Experiments like Cox et al. (2015), which employ relatively few questions, yield the narrowest range of expected costs for the three procedures.

To highlight the link between number of decision questions and cost, consider the impact of adding an $n + 1^{th}$ question to each procedure in turn. If the expected payment for question $n + 1$ is the same as the average for the original n questions, the expected cost to run a POR experiment would remain unchanged. In an RSP design, each subject would receive the same average payment, however the initial recruitment must increase slightly to account for the lower probability of making it through another question. PA experiments are the most sensitive to this addition, increasing the expected cost for all recruited subjects by that question's expected payout.

[119] As discussed above, and in section 1.3.3.

C One Task Treatment Sorting

Subjects were first sorted by their participant number into treatment arms. For the one task treatment, charities were initially placed into "tiers," each with two organizations, such that subjects would first be assigned to either of the first-tier charities, then once those reached a threshold, new participants were assigned to those in the second tier, etc. Although not ideal, this step was necessitated by uncertainty surrounding how many subjects could ultimately be recruited through the platform. The first "tier" included Feeding America and the Community Health Free Clinic, the second consisted of Doctors Without Borders USA and Our Companion Animal Sanctuary. The third tier (which was never reached) included Valley Outreach and the Animal Welfare Institute. Following the results of Eckel et al. (2018) who observed differences in dictator-game donations between local and national charities, early tiers were defined to include both a national-level and local level organization. (Unlike Eckel et al. (2018), this experiment did not include explicitly state-level organizations, however the distance between Ohio State's campus and the local organization was intentionally varied to attempt to induce similar psychological effects on participants.)

6

Appendices for Chapter 2: School Choice and Class Size Externalities

A Example Where All Students Gain

Consider the following five student, three school example. Assume that maximum capacities for each school are $|q_a| = |q_b| = 3$, $|q_c| = 2$.

Preference Ordering Held by Students								
i:	$a(1)$	$b(1)$	$c(1)$	$a(2)$	$b(2)$	$a(3)$	$b(3)$	$c(2)$
j:	$a(1)$	$b(1)$	$a(2)$	$b(2)$	$a(3)$	$b(3)$	$c(1)$	$c(2)$
k:	$a(1)$	$b(1)$	$a(2)$	$b(2)$	$a(3)$	$b(3)$	$c(1)$	$c(2)$
m:	$a(1)$	$a(2)$	$a(3)$	$b(1)$	$b(2)$	$b(3)$	$c(1)$	$c(2)$
p:	$a(1)$	$a(2)$	$a(3)$	$c(1)$	$c(2)$	$b(1)$	$b(2)$	$b(3)$
Priority Rankings for Each School								
a:	i	j	k	m	p			
b:	i	j	k	m	p			
c:	i	j	k	m	p			

Note that the above adjusted preference ranking corresponds with a "standard" problem where preferences are:

"Standard Problem" Preferences					
i:	a	b	c		
j:	a	b	c		
k:	a	b	c		
m:	a	b	c		
p:	a	c	b		
Priority Rankings for Each School					
a:	i	j	k	m	p
b:	i	j	k	m	p
c:	i	j	k	m	p

It is relatively straightforward to observe that the "standard" deferred acceptance algorithm would yield the resulting matching:

"Standard problem" DA outcome			
i:	$\underline{a}$	b	c
j:	$\underline{a}$	b	c
k:	$\underline{a}$	b	c
m:	a	$\underline{b}$	c
p:	a	$\underline{c}$	b

However, when this matching is translated into the extended problem (denoted by underlined outcomes in the table below), the deferred acceptance with voluntary withdrawals algorithm matching (denoted with boxes) strictly improves all student's outcomes:

Student outcomes for deferred acceptance algorithms.								
i:	$a(1)$	$b(1)$	$\boxed{c(1)}$	$a(2)$	$b(2)$	$\underline{a(3)}$	$b(3)$	$c(2)$
j:	$a(1)$	$b(1)$	$a(2)$	$\boxed{b(2)}$	$\underline{a(3)}$	$b(3)$	$c(1)$	$c(2)$
k:	$a(1)$	$b(1)$	$a(2)$	$\boxed{b(2)}$	$\underline{a(3)}$	$b(3)$	$c(1)$	$c(2)$
m:	$a(1)$	$\boxed{a(2)}$	$a(3)$	$\underline{b(1)}$	$b(2)$	$b(3)$	$c(1)$	$c(2)$
p:	$a(1)$	$\boxed{a(2)}$	$a(3)$	$\underline{c(1)}$	$c(2)$	$b(1)$	$b(2)$	$b(3)$

B Standard Deferred Acceptance Algorithm

For reference, the "standard" deferred acceptance algorithm is defined below. Note that this algorithm yields the Student Optimal Stable Matching (SOSM), and recall that it only incorporates preferences defined over schools, not school-class size pairs.

Round 1. *All students apply to their most preferred school. Every school a conditionally accepts the q_a students highest on a's priority ranking who applied (if the number of acceptable students who applied is greater than that school's maximum capacity), or all acceptable students if fewer than q_a students applied. Any student not offered a conditional acceptance is rejected from the school they applied to, eliminates that school from their preference list, and remains unmatched*

Round 2. *All unmatched students (i.e. those who are not holding a conditional acceptance at the start of round i) apply to their highest remaining preference option.*

Every school a conditionally accepts the highest-priority q_a students from among those who have applied this round and those previously holding conditional acceptances, and rejects the remaining students. All rejected students (including those who have had their conditional acceptance "revoked") strike that school from their preference ordering and remain unmatched.

The algorithm continues to repeat the procedure defined in round 2 from $2, \ldots, i$ until no students remain unmatched. Once that occurs, all students holding conditional acceptances are admitted to the corresponding schools.

C Algorithm Version 2: Simultaneous Proposals and Withdrawals

Beginning in some round t. Suppose the number of students currently holding seats at any given school a is $a(|q^t|)$. To clarify notation, class sizes with tildes ($a(\tilde{q})$) represent the contemporaneous number of seats filled (at any given step). A figure in round t without a tilde ($a(q)$) represents the start of round class size (at the start of step 1).

Step 1. ***Application Step:*** *All students who are not currently holding a seat from a school apply to their top remaining (i.e. "non-struck") school. Schools admit as many applicants as they have seats for. If the total number of students (including both those who have applied this round, and those currently holding seats at the school from having been admitted in previous rounds) is larger than the maximum capacity at the school, then admit the maximum possible number of students who are ranked as the top students in the priority ordering. All students currently holding offers do nothing this step.*

Step 2. ***Withdrawal Step:*** *This step applies to all students currently holding a seat at a school. Rank all students who were admitted during the previous step by their ordering in the school's priority list. Consider some school a. Beginning* *with the lowest ranked student (by a's priority ranking), compare each student i's preference profile ($\succ_i$) to the current school's class size ($\tilde{q}$), where $\tilde{q}$ includes not only the students who held a seat before Step 1, but also those who have not 'voluntarily withdrawn' as of the student currently being considered:*

1. *If there exists some other school $b \neq a$, such that the beginning of round class size at b (say r)*[120] *satisfies the property $b(r) \succ_i a(\tilde{q})$, then the student voluntarily withdraws him or herself from school a and becomes unmatched.*

2. *If there does not exist any other school $b \neq a$ that satisfies the aforementioned condition ($b(r) \succ_i a(\tilde{q})$), then the student does not withdraw and remains at school a.*

Step 3. ***Truncation Step:*** *If any student i withdrew from $a(\tilde{q})$ as a part of Step 2, then the entry for $a(q)$ is "struck off" the preference list for that student.*[121] *Similarly, for all students, remove all preference entries for school-class size pairs where the contemporaneous class size at the end of the round $\tilde{q}_a^t$ is at least one larger than the class size entry in the preference ranking.*

For an example of this second cause of truncation: suppose a student s_i is currently holding a seat at some school a and did not withdraw in the period. Listed after school a in s_i's preference list is some other school b, where the class size in the preference list is $b(x)$, but the current (start of step 3) class size of school b is $|\tilde{q}_b| \geq x + 1$ (that is, the current number of students sitting in school b's seats is at least one more than the maximum class size of interest for student i at that entry in their preference ranking). Why does this occur? Intuitively, we know that there are at least $x + 1$ students at school b at the beginning of step 3. Thus, there must be at least $x + 1$ students sufficiently happy at school b at the start of this round so as to have not withdrawn themselves in favor of some other school. (As will be addressed later), class sizes at schools can only weakly increase as the procedure continues, and thus the class size for school b will not shrink back down to size x, which is the

[120]Specifically, this refers to the class size at that alternative school immediately before step 1 for the current round occurred.

[121]I can add a specific example here if desired.

"requirement" for that entry in student s_i's preference list.[122]

C.1 Algorithm Version 2 Termination

First, consider the case where no student voluntarily withdraws from a seat at any point in the procedure. If no student withdraws, then the algorithm mirrors the standard deferred acceptance algorithm, which straightforwardly results in the DAwVW algorithm ending.

Attention is then turned to situations where at least one student withdraws from a school at some point during the algorithm. The only way for the algorithm not to terminate following such a withdrawal is if it initiates a cycle of applications and withdrawals.

What are the hypothetical cases where the algorithm will not terminate? Following the standard intuition of the deferred acceptance algorithms, as long as students get monotonically weakly worse off in every round (that is, they move down their preference rankings), the algorithm must eventually terminate (note that this says nothing about the stability or other properties of the matching finally reached, but merely that the procedure concludes).

Therefore, we focus on cases where cycles might occur in this alternative withdrawal approach. At first glance, it appears that this might be problematic- if a chain of withdrawals from other students at other schools leads to some school a's class size to be reduced, there may very well be some other student who now prefers this smaller school a to their current seat. If so, they withdraw, take their place at a, leaving their old school's class size now smaller. If these withdrawals and applications lead to a cycle, then the algorithm will never terminate.

What must be shown, then, is that these cycles are actually not permitted under the current framework. To do so, we first specifically consider a generalized case of the necessary conditions for this to occur.

[122]Note that this does not imply that student s_i will never apply to school b, only that they won't do so on the condition of the class size being x.

Consider some student s_i, and suppose the algorithm is in progress at stage $t-1$. Suppose at the beginning of $t-1$ student s_i was unmatched (that is, was not temporarily holding a seat at any school), and chose to apply to some school c_a with a beginning of round $t-1$ class size of $|q_a^{t-1}|$. We know, by the nature of s_i applying to c_a in round $t-1$ that there exists no other school $c_b \neq c_a$, with start of period $t-1$ class size $|q_b^{t-1}|$ such that $c_b(|q_b^{t-1}|) \succ_{s_i} c_a(|q_a^{t-1}|)$, otherwise student s_i would have chosen to apply to c_b at the beginning of the round instead of c_a.[123] The question then becomes does there exist a case where student i would withdraw from school c_a with the class size of $|q_a^{t-1}|$ in some round after $t-1$?

Remark 1. *Note that this is ultimately the necessary condition for a cycle to occur- if school c_a grows larger than $|q_a^{t-1}|$, and that's what induces the withdrawal, then student s_i is still moving down their preference ranking over rounds- they're withdrawing because in some round $t > t-1$, school c_a grew, and there existed some other school c_b such that $c_a(|q_a^{t-1}|) \succ_{s_i} c_b(|q_b^t|) \succ_{s_i} c_a(|q_a^t|)$. Ultimately, they're moving from what they applied to originally ($c_a(|q_a^{t-1}|)$) to some alternative ($c_b(|q_b^t|)$) because the school they had originally applied to has now grown to ($c_a(|q_a^t|)$). From the preference ranking profile, they're still moving from the first option on the list mentioned here to the second option, instead of being "forced" by the system into accepting the worst of the three.*

So, when would student s_i withdraw from c_a with class size $|q_a^{t-1}|$ after round $t-1$? Only if, in some period $t > t-1$ there existed some school $c_b \neq c_a$ that is higher on s_i's preference list than $c_a(|q_a^{t-1}|)$. Since there was no such school at the start of the earlier round, this also necessarily implies:

$$c_b(|q_b^t|) \succ_{s_i} c_b(|q_b^{t-1}|) \qquad \text{(Condition 1)}$$

Because we assume preferences in class sizes are monotonic, we know Condition 1 implies that $|q_b^t| < |q_b^{t-1}|$. Thus, for some other school to become "sponta-

[123] All this is saying is that, at the start of round $t-1$, given all the current class sizes of schools, school c_a must have been the most preferred available option for student s_i, otherwise they would have chosen to apply somewhere else.

neously" more preferred to c_a in some round t it must be the case that school c_b shrank in the period before t.

However, this condition itself only occurs if some student from c_b him or herself chose to withdraw from c_b in the round before t without an increase in class size. In general, a class size from one school can only shrink in round x if there exists some other school that shrank in the round preceding x.

Thus, we only must show that there exists a period where class sizes did not shrink, since this would directly imply that no subsequent rounds can observe any school shrinking in size. In fact, this condition holds in round 1, where all school class sizes are zero, and there exists at least one student who is tentatively matched at the end of round 1 to some school. Thus, since in round 1 class sizes weakly grow, there exists no school that shrinks in the period before round 2, and thus no school can shrink in round 2. This continues until round t, disproving the initial assumption that such a cycle could occur.